MASTER EN SCIENCES MATHÉMATIQUES

Sur l'usage des fonctions de Bregman dans un algorithme de point proximal non linéaire. Applications à la programmation convexe.

BROTCORNE, Luce

Award date:
1992

Awarding institution:
Universite de Namur

FACULTES UNIVERSITAIRES NOTRE-DAME DE LA PAIX
NAMUR
FACULTE DES SCIENCES

SUR L'USAGE DES FONCTIONS DE BREGMAN DANS UN ALGORITHME DE POINT PROXIMAL NON LINEAIRE. APPLICATIONS A LA PROGRAMMATION CONVEXE.

Promoteur : J.-J. STRODIOT

Luce BROTCORNE

Année académique : 1991-1992

J'exprime toute ma gratitude envers Monsieur le Professeur Jean-Jacques STRODIOT pour son aide et sa guidance qui me furent d'un très grand apport lors de la réalisation de ce mémoire.

Je tiens également à remercier ma famille et mes amis pour leur soutien et leurs encouragements tout au long de mes études.

Chapitre 1

Introduction

La méthode du point proximal a été introduite par Martinet [15] pour *rechercher le minimum d'une fonction convexe, propre et semi-continue inférieurement sur l'espace euclidien* $I\!R^n$. L'algorithme associé à cette méthode génère une suite $\{x^k\}$ telle que

$$x^{k+1} = \text{argmin}_{\,x\in I\!R^n} \{f(x) + \frac{1}{2\lambda_k}||x - x^k||^2\} \tag{1.1}$$

où $\{\lambda_k\}$ est une suite de scalaires strictement positifs.

Ce problème de minimisation est, en réalité, un cas particulier de la recherche d'un zéro d'un opérateur maximal monotone T où T est le sous-différentiel de la fonction f. Dans son article [20], Rockafellar a étudié le problème général de la *recherche d'un zéro d'un opérateur maximal monotone* T *quelconque.* L'algorithme relatif génère une suite $\{x^k\}$ telle que

$$x^{k+1} = (I + \lambda_k T)^{-1}(x^k) \tag{1.2}$$

où $\{\lambda_k\}$ est une suite de scalaires strictement positifs.

Dans ce mémoire, nous généraliserons la méthode du point proximal en remplaçant le terme quadratique ajouté à la fonction f par un terme plus général de la forme $D_h(x, x^k)$ où D_h est une "distance" associée à une fonction h appelée *fonction de Bregman.* Dans ce cas, la suite $\{x^k\}$ est définie par l'itération :

$$x^{k+1} = \text{argmin}_{\,x\in I\!R^n} \{f(x) + \frac{1}{\lambda_k} D_h(x, x^k)\} \tag{1.3}$$

où $\{\lambda_k\}$ est une suite de scalaires strictement positifs.

Comme dans le cas classique, nous ferons appel à la théorie des opérateurs pour analyser la convergence de cet algorithme.

Dans le chapitre 2, nous généraliserons l'algorithme du point proximal de Rockafellar. Le but du nouvel algorithme consiste toujours à rechercher un zéro d'un opérateur maximal monotone T quelconque, mais les itérations font intervenir une fonction de Bregman h. Plus précisément,

$$x^{k+1} = ((\nabla h + \lambda_k T)^{-1} \circ \nabla h)(x^k) \tag{1.4}$$

où $\{\lambda_k\}$ est une suite de scalaires strictement positifs.
Cet algorithme, appelé *algorithme du point proximal non linéaire utilisant des fonctions de Bregman*, est dû à Eckstein [10]. Lorsque la fonction h est définie par $h(x) = \frac{1}{2}||x||^2$, nous retrouvons l'algorithme de Rockafellar (itération (1.2)).

Ensuite, dans le chapitre 3, nous appliquerons l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman au problème de minimisation convexe. Nous retrouverons ainsi l'itération proximale non linéaire (1.3).
De plus, nous remarquerons que sur des architectures de calcul en parallèle, la résolution d'un problème linéaire par *la méthode du point proximal non linéaire* utilisant la fonction de Bregman $h(x) = -entx$ (où $entx$ désigne la fonction entropique), pourra être plus rapide que celle obtenue par une méthode directe (comme l'algorithme du simplexe ou les méthodes de points intérieurs). Cette méthode de résolution d'un problème linéaire fut suggérée par Eriksson [11].

Dans la seconde partie de ce mémoire, nous définirons une *méthode des multiplicateurs utilisant des fonctions de Bregman.* Pour des raisons de clarté, nous envisagerons séparément les problèmes de minimisation convexe avec contraintes d'égalités (PE) ou d'inégalités (PI).

Dans le chapitre 4, nous appliquerons l'algorithme du point proximal non linéaire à la formulation duale lagrangienne d'un problème de programmation convexe (PE) ou (PI). L'opérateur T sera cette fois identifié au sous-différentiel de l'opposé de la fonction duale. Nous obtiendrons ainsi une nouvelle classe de méthodes non quadratiques des multiplicateurs. Nous remarquerons que par les choix $h(x) = \frac{1}{2}||x||^2$ et $h(x) = -entx$ nous retrouvons respectivement, la méthode du lagrangien augmenté classique et la méthode exponentielle des multiplicateurs.

Ensuite, dans le chapitre 5, nous appliquerons l'algorithme du point proximal non linéaire à la formulation "primal-dual" d'un problème de minimisation convexe (PI). Nous obtiendrons ainsi la première méthode proximale non quadratique des multiplicateurs. Nous vérifierons que le choix particulier $h(x) = \frac{1}{2}||x||^2$ nous donne à nouveau la méthode proximale des multiplicateurs étudiée par Rockafellar [17].

Pour terminer, nous présenterons dans le chapitre 6 les difficultés rencontrées lors de l'application de l'algorithme du point proximal non linéaire à *la recherche d'un zéro d'une somme de plusieurs opérateurs maximaux monotones* (qui n'est en général pas maximale monotone).

La référence principale de ce mémoire est [10], "Non Linear Proximal Point algorithms using Bregman Functions, with Applications to Convex Programming", J. Eckstein.

Chapitre 2

Recherche d'un zéro d'un opérateur maximal monotone

2.1 Motivation

La motivation principale de ce chapitre est l'étude du problème d'optimisation convexe suivant :

$$(\mathrm{P}_1) \qquad \begin{cases} \text{Minimiser} & f(x) \\ \text{s.c.} & x \in X \end{cases}$$

où f est une fonction propre, convexe, semi-continue inférieurement définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$ et X est un ensemble convexe, fermé, non vide de $I\!R^n$. De plus, pour le bien fondé du problème, nous supposerons que l'ensemble dom $f \cap X$ est non vide. Pour le rappel des définitions concernant ce problème, on pourra se référer aux annexes I et III.

Il est immédiat que le problème (P_1) est équivalent à :

$$(\mathrm{P}_2) \qquad \text{Minimiser } f(x) + \delta_X(x) \quad \text{sur tout } x \in I\!R^n$$

où $\delta_X(.)$ désigne la fonction indicatrice de l'ensemble X. C'est une fonction propre, convexe (cfr exemple A.I.14), définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$ par

$$\delta_X(x) = \begin{cases} 0 & \text{si } x \in X \\ +\infty & \text{sinon} \end{cases}$$

De plus, X étant un ensemble fermé, cette fonction $\delta_X(.)$ est également s.c.i. (cfr exemple A.III.5).

Par ailleurs, si nous définissons la fonction f_X de $I\!R^n$ dans $I\!R \cup \{+\infty\}$ par $f_X := f + \delta_X$, le problème (P_2) peut se réécrire sous la forme équivalente :

(P_3) Minimiser $f_X(x)$ sur tout $x \in I\!R^n$

où la fonction f_X, somme de deux fonctions propres, convexes, et s.c.i., est elle aussi propre convexe et s.c.i.

Ecrivons à présent la condition d'optimalité de (P_3) :

$$0 \in \partial(f_X)(x)$$

où $\partial(f_X)(x)$ désigne le *sous-différentiel* de la fonction convexe f_X au point x, c'est-à-dire,

$$\partial f_X(x) = \{y \in I\!R^n \text{ tel que } f_X(z) \geq f_X(x) + < y, z - x > \ \forall\, z \in I\!R^n\} .$$

De plus, comme nous le verrons dans le paragraphe 2.2 (exemple 2.5), la fonction $f_X(.)$ étant propre, convexe et s.c.i., son sous-différentiel $\partial f_X(.)$ est un opérateur maximal monotone.

La résolution du problème (P_1) est donc équivalente à la recherche d'un zéro de l'opérateur maximal monotone $\partial f_X(.)$. Plus précisément, le problème (P_1) peut se réécrire sous la forme suivante :

(P_4) Trouver $x \in I\!R^n$ tel que $0 \in \partial f_X(x)$.

Ce chapitre sera consacré à la recherche d'un zéro d'un opérateur maximal monotone quelconque T défini sur $I\!R^n$.

Nous commencerons donc par présenter des résultats essentiels concernant les opérateurs monotones, maximaux monotones et cycliquement monotones (cfr paragraphe 2.2).

Ensuite, nous analyserons brièvement en quoi consiste l'algorithme du point proximal proposé par Rockafellar (cfr paragraphe 2.3).

Enfin, nous étudierons une généralisation utilisant des fonctions de Bregman de cet algorithme. Cette généralisation est due à Eckstein (cfr paragraphe 2.4).

Par la suite, dans le chapitre 3, nous particulariserons les résultats obtenus dans le cadre général au problème (P_1), motivation particulière de notre étude.

2.2 Les opérateurs monotones, maximaux monotones, cycliquement monotones

Dans ce paragraphe, nous rappelons quelques résultats importants concernant les opérateurs monotones, maximaux monotones et cycliquement monotones. Ces résultats peuvent se trouver dans les travaux de Brézis [2].

Soit l'espace $I\!R^n$ muni du produit scalaire $< .,. >$ et de la norme euclidienne usuelle $||.||$. Etant donné $D \subseteq I\!R^n$, on désignera par $\overline{D}$, int D et conv D respectivement la fermeture, l'intérieur et l'enveloppe convexe de D dans $I\!R^n$.
Si X est un convexe fermé de $I\!R^n$, alors $\text{Proj}_X\, x$ désigne la projection de x sur X.

Définition 2.1 :
Un *opérateur* T sur $I\!R^n$ est une application de $I\!R^n$ dans $P(I\!R^n)$, où $P(I\!R^n)$ désigne l'ensemble des parties de $I\!R^n$.
Son *domaine* est l'ensemble

$$D(T) = \{x \in I\!R^n \text{ tel que } T(x) \neq \emptyset\}$$

Son *image* est l'ensemble

$$R(T) = U_{x \in I\!R^n}\, T(x)$$

Son *graphe* est l'ensemble

$$G(T) = \{(x, y) \in I\!R^n \times I\!R^n \text{ tel que } y \in T(x)\}$$

Notons que par la suite nous identifierons l'opérateur T à son graphe.

Définition 2.2 :
L'opérateur T est *univoque* si, pour tout x dans $I\!R^n$, l'ensemble $T(x)$ contient au plus un élément. Il est dit *multivoque* sinon.

Définition 2.3 :
Soient A et B deux opérateurs sur $\mathbb{R}^n$ et $\lambda, \mu \in \mathbb{R}$. Alors $\lambda A + \mu B$ est l'opérateur qui à un point x dans $\mathbb{R}^n$ associe

$$(\lambda A + \mu B)(x) = \{\lambda u + \mu v \text{ où } u \in A(x) \text{ et } v \in B(x)\} .$$

Notons que $D(\lambda A + \mu B) = D(A) \cap D(B)$.

Définition 2.4 :
L'opérateur *inverse* de T, c'est-à-dire T^{-1} est l'opérateur dont le graphe est le symétrique du graphe de T, autrement dit,

$$y \in T^{-1}(x) \Leftrightarrow x \in T(y) .$$

Nous en déduisons immédiatement que $D(T^{-1}) = R(T)$.

Définition 2.5 :
Un opérateur T sur $\mathbb{R}^n$ est une *contraction* si

$$\forall\, x_1, x_2 \in D(T),\ \forall\, y_1 \in T(x_1),\ \forall\, y \in T(x_2) \quad ||y_1 - y_2|| \leq ||x_1 - x_2||$$

Un tel opérateur est toujours univoque.

Définition 2.6 :
L'ensemble des opérateurs sur $\mathbb{R}^n$ est ordonné par *l'inclusion des graphes*, c'est-à-dire

$$A \subseteq B \Leftrightarrow A(x) \subseteq B(x) \quad \text{pour tout } x \text{ dans } \mathbb{R}^n$$

où A et B sont deux opérateurs sur $\mathbb{R}^n$.

Définition 2.7 :
Un opérateur T sur $\mathbb{R}^n$ est dit *monotone* si

$$\forall\, x_1, x_2 \in D(T),\ \forall\, y_1 \in T(x_1),\ \forall\, y_2 \in T(x_2) \ < y_1 - y_2, x_1 - x_2 > \geq 0$$

Voici à présent quelques exemples d'opérateurs monotones.

Exemple 2.1 :

- Soit T un opérateur monotone sur $I\!R^n$, alors les opérateurs $T^{-1}, \lambda T$ pour tout $\lambda \geq 0$, et $\overline{T}$ (fermeture de T dans $I\!R^n \times I\!R^n$) sont monotones.
- Soient A et B deux opérateurs monotones, alors $A + B$ est monotone.
- Soit J une contraction de $D \subseteq I\!R^n$ dans $I\!R^n$, alors l'opérateur $I - J$ est monotone.
- Soit X un ensemble convexe fermé, alors l'opérateur $x \rightarrow \mathrm{Proj}_X x$ est monotone.

Exemple 2.2 :
Soit φ une fonction propre et convexe sur $I\!R^n$. Alors le sous-différentiel de φ, noté $\partial\varphi$, est un opérateur monotone sur $I\!R^n$.

Preuve :
Soient $x_1, x_2 \in I\!R^n, y_1 \in \partial\varphi(x_1)$ et $y_2 \in \partial\varphi(x_2)$.
Il suit de la définition du sous-différentiel que :

$$\varphi(x_2) \geq \varphi(x_1) + < y_1, x_2 - x_1 >$$

et

$$\varphi(x_1) \geq \varphi(x_2) + < y_2, x_1 - x_2 > \ .$$

Dès lors,

$$\varphi(x_2) + \varphi(x_1) \geq \varphi(x_1) + < y_1, x_2 - x_1 > + \varphi(x_2) + < y_2, x_1 - x_2 >$$

d'où, après simplification,

$$0 \geq < y_1 - y_2, x_2 - x_1 >$$

ou encore

$$0 \leq < y_2 - y_1, x_2 - x_1 > \ .$$

Définition 2.8 :
Un opérateur T sur $I\!R^n$ est dit *strictement monotone* si

$$\forall\ x_1, x_2 \in D(T),\ \forall\ y_1 \in T(x_1),\ \forall\ y_2 \in T(x_2)\ < y_1 - y_2, x_1 - x_2 > > 0\ .$$

■

Exemple 2.3 :
Soit φ une fonction propre et strictement convexe sur $I\!R^n$. Alors le sous-différentiel de φ, noté $\partial\varphi$, est un opérateur strictement monotone sur $I\!R^n$.

Enonçons maintenant une caractérisation des opérateurs monotones.

Proposition 2.1 :
Soit T un opérateur sur $I\!R^n$. T est monotone si et seulement si

$$\forall\, x_1, x_2 \in D(T),\ \forall\, y_1 \in T(x_1),\ \forall\, y_2 \in T(x_2),\ \forall\, \lambda > 0 ||x_1 - x_2|| \leq ||(x_1 - x_2) + \lambda(y_1 - y_2)||$$

Preuve : cfr [2], proposition 2.1, p. 21.

Cette condition exprime que pour tout $\lambda > 0$, l'opérateur $(I + \lambda T)^{-1}$ est une contraction de $R(I + \lambda T)$ dans $I\!R^n$. Autrement dit, pour tout y dans $I\!R^n$, l'équation $y \in x + \lambda T(x)$ admet *au plus* une solution.

Nous allons à présent considérer une classe d'opérateurs pour lesquels l'équation $y \in x + \lambda T(x)$ admet *exactement* une solution x pour tout y dans $I\!R^n$ et pour tout $\lambda > 0$. Il s'agit des opérateurs maximaux monotones.

Définition 2.9 :
Un opérateur T sur $I\!R^n$ est *maximal monotone* si son graphe $G(T)$ n'est pas inclus de façon stricte dans le graphe d'un autre opérateur monotone T' sur $I\!R^n$.
Autrement dit, T est *maximal monotone* si et seulement si T est monotone et si

$$\forall\, (x, y) \in I\!R^n \times I\!R^n \text{ tel que } \forall (\xi, \eta) \in T \ < y - \eta, x - \xi > \geq 0 \text{ , alors } y \in T(x)$$

Citons à présent quelques exemples d'opérateurs maximaux monotones.

Exemple 2.4 :
Soit T un opérateur maximal monotone. Alors les opérateurs T^{-1} et λT pour tout $\lambda > 0$ sont maximaux monotones.

Exemple 2.5 :
Soit φ une fonction propre et convexe sur $I\!R^n$. Si φ est s.c.i., alors le sous-différentiel de φ, noté $\partial\varphi$, est un opérateur maximal monotone.

Indiquons maintenant deux propriétés des opérateurs maximaux monotones.

Proposition 2.2 :
Soit T un opérateur sur $\mathbb{R}^n$. Les trois propriétés suivantes sont équivalentes :

1. T est maximal monotone
2. T est monotone et $R(I+T) = \mathbb{R}^n$
3. Pour tout $\lambda > 0, (I + \lambda T)^{-1}$ est une contraction définie sur l'espace $\mathbb{R}^n$ tout entier.

Preuve : cfr [2], proposition 2.2, p. 23.

Proposition 2.3 :
Soit T un opérateur maximal monotone sur $\mathbb{R}^n$. Alors $\overline{D(T)}$ est un ensemble convexe.

Preuve : cfr [2], théorème 2.2, p. 27.

La fermeture d'un opérateur maximal monotone T étant monotone il suit que, T est fermé dans $\mathbb{R}^n \times \mathbb{R}^n$. Plus précisément, on dispose de la proposition suivante :

Proposition 2.4 :
Soit T un opérateur maximal monotone sur $\mathbb{R}^n$. Considérons deux suites $\{x_n\}$ et $\{y_n\}$ telles que $\forall\, n \in \mathbb{N} \quad y_n \in T(x_n)$.
Si $x_n \to x$ et $y_n \to y$, alors $y \in T(x)$.

Preuve : cfr [2], proposition 2.5, p. 27.

Remarquons également que la somme de deux opérateurs maximaux monotones A et B n'est pas nécessairement maximale monotone. En effet, le domaine de cet opérateur somme, ie $D(A+B) = D(A) \cap D(B)$, peut être vide. C'est pourquoi, on a établi des conditions suffisantes pour que la somme de deux opérateurs maximaux monotones soit maximale monotone. Citons les deux propriétés suivantes :

Théorème 2.1 :
Soient A et B deux opérateurs maximaux monotones. Si (Int $D(A)) \cap D(B) \neq \emptyset$, alors $A+B$ est maximal monotone.
De plus, $\overline{D(A) \cap D(B)} = \overline{D(A)} \cap \overline{D(B)}$.

Preuve : cfr [2] corollaire 2.7, p. 36.

Proposition 2.5 :
Soient A un opérateur maximal monotone et B un opérateur monotone, lipschitzien. Alors $A + B$ est maximal monotone.

Preuve : cfr [2], lemme 2.4, p. 34.

Pour terminer, intéressons-nous à une autre classe d'opérateurs monotones, en l'occurrence les opérateurs cycliquement monotones.

Définition 2.10 :
Un opérateur T sur $I\!R^n$ est *cycliquement monotone* si pour toute suite cyclique $\{x_k\}$ de $D(T)$, ie $x_0, x_1, \cdots x_n = x_0$, et toute suite $\{y_k\}$ telle que $y_i \in T(x_i)$ $i = 1, 2 \cdots n$, on a

$$\sum_{i=1}^{n} < x_i - x_{i-1}, y_i > \geq 0 \; .$$

Notons que tout opérateur cycliquement monotone est évidemment monotone.

Exemple 2.6 :
Soit φ une fonction propre et convexe sur $I\!R^n$. Alors le sous-différentiel de φ, noté $\partial\varphi$, est un opérateur cycliquement monotone.
En particulier, $\partial\varphi$ est *trimonotone* (c'est-à-dire cycliquement monotone de cycle 3).

Preuve :
Soient $x_1, x_2, \cdots x_n = x_0$ et $y_i \in \partial\varphi(x_i)$ $i = 1, \cdots, n$.
φ étant propre, nous obtenons que $\varphi(x_i) < \infty$ pour tout x_i.
De plus, par la définition du sous-différentiel de φ,

$$\varphi(x_{i-1}) - \varphi(x_i) \geq < x_{i-1} - x_i, y_i > \quad i = 1, \cdots, n \; .$$

Dès lors,

$$\sum_{i=1}^{n} < x_{i-1} - x_i, y_i > \leq \sum_{i=1}^{n} (\varphi(x_{i-1}) - \varphi(x_i)) \; ,$$

c'est-à-dire, puisque la suite $\{x_i\}$ est cyclique,

$$\sum_{i=1}^{n} < x_{i-1} - x_i, y_i > \leq 0 \; ,$$

ou encore

$$\sum_{i=1}^{n} < x_i - x_{i-1}, y_i > \geq 0 \ .$$

Par conséquent l'opérateur $\partial\varphi$ est cycliquement monotone. Considérant le cas particulier $n = 3$, nous déduisons également que $\partial\varphi$ est trimonotone. ■

Dans les paragraphes suivants, nous ferons des références explicites à ces résultats.

2.3 L'algorithme du point proximal de Rockafellar

L'algorithme du point proximal de Rockafellar est destiné à la recherche d'un zéro d'un opérateur maximal monotone.

Soit T un opérateur maximal monotone défini sur $I\!R^n$.
Puisque T est maximal monotone il suit de la proposition 2.2 que, pour tout réel $\lambda > 0$, l'opérateur univoque $P := (I + \lambda T)^{-1}$ est une contraction définie sur l'espace $I\!R^n$ tout entier et à valeurs dans $I\!R^n$.
Cette propriété permet de ramener le problème de la recherche d'un zéro de T à celui d'un point fixe de P. En effet, on a successivement :

$$\begin{aligned} P(z) = z \quad &\Leftrightarrow \quad z \in (I + \lambda T)z \quad \forall\, \lambda > 0 \\ &\Leftrightarrow \quad 0 \in T(z) \ . \end{aligned}$$

L'algorithme du point proximal de Rockafellar procède comme suit : étant donné un point z^0 dans $I\!R^n$, on engendre une suite $\{z^k\}$ définie par

$$z^{k+1} = P_k(z_k)$$

où $P_k := (I + \lambda_k T)^{-1}$ et $\{\lambda_k\}$ est une suite de réels strictement positifs.

Le théorème suivant établit la convergence de cet algorithme.

Théorème 2.2 :
Soit $\{z^k\}$ une suite générée par l'algorithme du point proximal où la suite $\{\lambda_k\}$ a été choisie minorée par $\overline{\lambda} > 0$.
Supposons que la suite $\{z^k\}$ soit bornée (cela est vérifié, si et seulement si, il existe un zéro

de l'opérateur T). Alors, sous les hypothèses précédentes, la suite $\{z^k\}$ converge vers un point z tel que

$$0 \in T(z) .$$

De plus

$$\lim_{k \to +\infty} ||z^{k+1} - z^k|| = 0 = \lim_{k \to +\infty} ||Q_k(z^k)\,|| \text{ où } Q_k = I - P_k .$$

Rockafellar a également étudié la vitesse de convergence de cet algorithme. De façon générale, on peut dire que plus vite la suite $\{\lambda^k\}$ croit, plus rapide est la convergence (cfr [20], théorèmes 2 et 3).

2.4 L'algorithme du point proximal non linéaire utilisant des fonctions de Bregman

Dans ce paragraphe, nous désirons généraliser l'algorithme du point proximal de Rockafellar. Le but du nouvel algorithme consiste toujours à rechercher un zéro d'un opérateur maximal monotone T, mais les itérations feront intervenir une fonction h appelée fonction de Bregman. Pour un choix particulier de h, on retrouvera l'algorithme de Rockafellar.

Dans un premier temps, nous donnerons une définition précise d'une fonction de Bregman.

Ensuite, nous montrerons que la recherche d'un zéro d'un opérateur maximal monotone T sur $I\!R^n$ est équivalente à la recherche d'un point fixe de l'opérateur $P = (\nabla h + \lambda T)^{-1} \circ \nabla h$ où h est une fonction de Bregman et λ un scalaire strictement positif. Par analogie avec l'algorithme de Rockafellar, nous serons alors amenés à définir l'itération de l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman par

$$x^{k+1} = ((\nabla h + \lambda_k T)^{-1} \circ \nabla h)\ (x^k)$$

où $\{\lambda_k\}$ est une suite de scalaires strictement positifs.

Enfin, nous proposerons des conditions d'existence des itérés et nous analyserons la convergence de cet algorithme.

2.4.1 Fonctions de Bregman

Soient $X \subseteq I\!R^n$ un ensemble convexe, fermé, non vide et $y \in I\!R^n \setminus X$. Le point de X le plus proche de y est le point unique x^* résolvant le problème convexe suivant :

$$d(x^*, y) = \inf_{x \in X} \ \{d(x, y) = ||x - y||\}$$

où $||.||$ désigne la norme euclidienne usuelle sur $I\!R^n$. Le point x^*, appelé *la projection de* y *sur* X, est noté $x^* = P_X(y)$.

Nous remplaçons à présent la distance euclidienne classique par *une nouvelle mesure de distance* "$D(x, y)$", satisfaisant la relation suivante :

$$\text{``}D(x,y)\text{''} \begin{cases} = 0 & \text{si } x = y \\ > 0 & \text{sinon.} \end{cases}$$

Soit $S \subseteq I\!R^n$ un ensemble ouvert non vide. Une telle mesure de distance peut être engendrée par une fonction $h : \overline{S} \to I\!R^n$ différentiable sur S, continue et strictement convexe sur $\overline{S}$. En effet, définissons la fonction D_h de $\overline{S} \times S$ dans $I\!R$ par

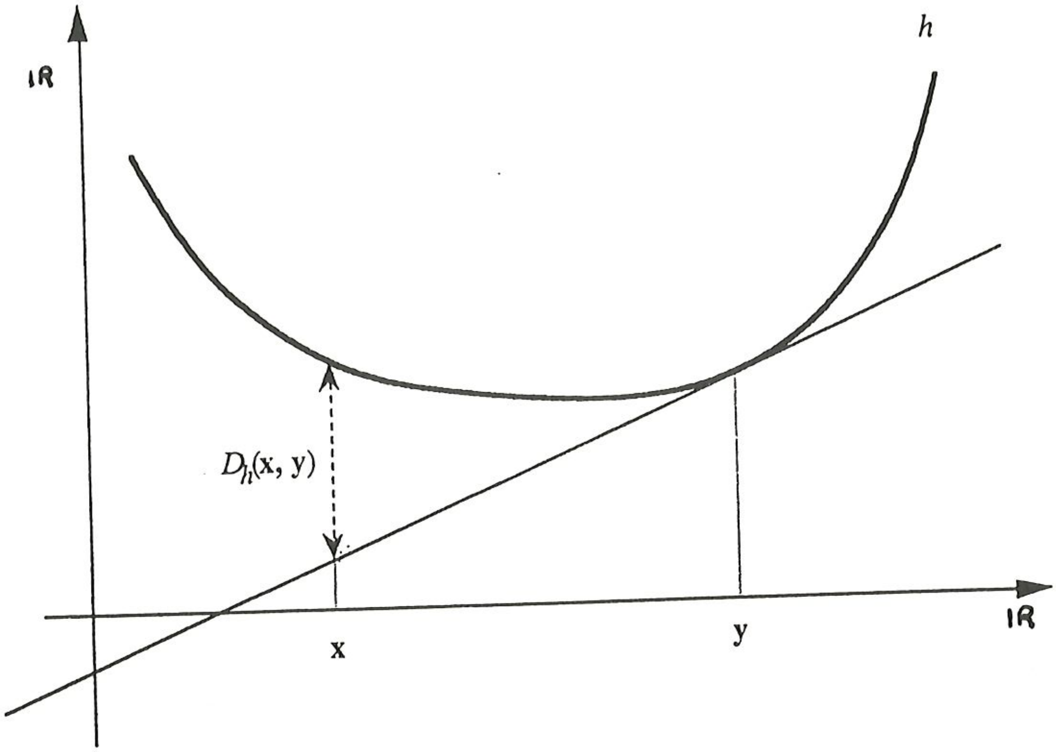
$$D_h(x, y) = h(x) - h(y) - < \nabla h(y), x - y > \ . \tag{2.1}$$

Comme la fonction h est strictement convexe, nous obtenons immédiatement que

$$D_h(x, y) \geq 0 \ \ \forall\, x \in \overline{S} \ \forall\, y \in S \tag{2.2}$$

$$\text{et} \ \ D_h(x, y) = 0 \ \Leftrightarrow x = y \tag{2.3}$$

D'un point de vue géométrique, $D_h(x, y)$ peut être interprétée comme *la distance entre* x *et le plan tangent à la fonction* h *en* y :



Notons que la fonction D_h ne définit pas une métrique. En effet, en général, D_h ne vérifie ni la propriété de symétrie, ni l'inégalité triangulaire.

Nous allons à présent imposer quelques hypothèses supplémentaires sur la fonction h, de sorte que la distance $D_h(.,y)$ définie par (2.1) se comporte sur $\overline{S}$ de façon semblable à $\frac{1}{2}||. - y||^2$ sur $I\!R^n$. A cet effet, nous définissons une fonction de Bregman de la façon suivante :

Définition 2.11 : Fonction de Bregman
Soit S un sous-ensemble convexe ouvert de $I\!R^n$. Une fonction $h : \overline{S} \to I\!R$ est appelée *fonction de Bregman de zone* S si elle vérifie les propriétés citées ci-dessous :

(i) h est continûment différentiable sur S

(ii) h est strictement convexe sur $\overline{S}$

(iii) h est continue sur $\overline{S}$

(iv) $\forall\ \alpha \in I\!R$, les ensembles de niveau $L_1(y,\alpha) := \{x \in \overline{S}$ tel que $D_h(x,y) \leq \alpha\}$ sont bornés pour tout $y \in S$.

$\forall\ \beta \in I\!R$, les ensembles de niveau $L_2(x, \beta) := \{y \in S$ tel que $D_h(x,y) \leq \beta\}$ sont bornés pour tout $x \in \overline{S}$.

(v) Soit une suite $\{y^k\}$ contenue dans S telle que $\lim_{k\to\infty}\ y^k = y^*$.
Alors $\lim_{k\to\infty}\ D_h(y^*, y^k) = 0$

(vi) Soient $\{x^k\}$ et $\{y^k\}$ deux suites contenues dans S telles que $\{x^k\}$ est bornée, $\lim_{k\to\infty}\ y^k = y^* \in \overline{S}$ et $\lim_{k\to\infty}\ D_h(x^k, y^k) = 0$. Alors $\lim_{k\to\infty}\ x^k = y^*$.

Constatons tout d'abord que toute fonction de Bregman h définie initialement sur $\overline{S}$ peut être étendue à $I\!R^n$ en posant $h(x) = +\infty\ \forall\ x \notin \overline{S}$.
En vertu de [19, p. 24 et p. 56], cette extension de h est une fonction propre, convexe et fermée.

D'autre part, par analogie avec l'opérateur de projection usuel sur un ensemble convexe, fermé, non vide X, nous définissons *l'opérateur de D_h - projection* de y sur X de la façon suivante :

$$P_{h,X}(y) := \text{argmin}\ \{D_h(x,y) \text{ tel que } x \in \bar{S} \cap X\}\ , \tag{2.4}$$

où le membre de droite de l'égalité désigne l'ensemble des minima de la fonction $D_h(.,y)$ sous contraintes $x \in \bar{S} \cap X$ et où

$$D_h(x,y) = h(x) - h(y) - < \nabla h(y), x - y >\ . \tag{2.5}$$

Avant de considérer quelques exemples de fonction de Bregman, il nous semble utile d'établir les trois théorèmes suivants :

<u>**Théorème 2.3 :**</u>
Soit $h : I\!R^n \to I\!R$. Si h vérifie les deux conditions suivantes :

1. h est une fonction deux fois continûment différentiable et strictement convexe sur $I\!R^n$.
2. $\lim_{||x||\to\infty} \frac{h(x)}{||x||} = +\infty$

alors h est une fonction de Bregman de zone $I\!R^n$.

<u>**Preuve :**</u> cfr [4], théorème 5.1, p. 435.

Théorème 2.4 :
Si h est une fonction de Bregman de zone $S \subseteq I\!R^n$, alors la fonction g définie par $g(x) = h(x)+ < c, x > +b$ où $c \in I\!R^n$ et $b \in I\!R$, est également une fonction de Bregman de zone S.
De plus, $\forall\ (x, y) \in \overline{S} \times S \quad D_g(x, y) = D_h(x, y)$.

Preuve :

Puisque la fonction affine f, définie de $I\!R^n$ dans $I\!R$ par $f(x) =< c, x > +b$, est continûment différentiable sur $I\!R^n$ et que la fonction h est une fonction de Bregman de zone S, il est immédiat que la fonction g est continûment différentiable sur S, continue et strictement convexe sur $\bar{S}$.

D'autre part, la fonction g induit la même fonction distance que la fonction de Bregman h.
En effet, par définition, $\forall\ x \in \bar{S}$ et $\forall\ y \in S \quad D_g(x, y) = g(x) - g(y)- < \nabla g(y), x - y >$.
Or $\nabla g(y) = \nabla h(y) + c$, dès lors,

$$D_g(x, y) = h(x)+ < c, x > -b - h(y)- < c, y > +b- < \nabla h(y) + c, x - y > \ .$$

D'où après simplification,

$$D_g(x, y) = h(x) - h(y)- < \nabla h(y), x - y > ,$$

autrement dit $\quad D_g(x, y) = D_h(x, y)$.

La thèse résulte alors des propriétés de la distance D_h associée à la fonction de Bregman h. ■

Théorème 2.5 :
Soient h^1 et h^2 deux fonctions de Bregman de zone respective $S^1 \subseteq I\!R^n$ et $S^2 \subseteq I\!R^m$.
Alors la fonction h définie de $\overline{S^1} \times \overline{S^2}$ dans $I\!R$ par $h(x, p) = h^1(x) + h^2(p)$ est une fonction de Bregman de zone $S = S^1 \times S^2$.
De plus

$$D_h((x_1, p_1), (x_2, p_2)) = D_{h^1}(x_1, x_2) + D_{h^2}(p_1, p_2)$$

pour tout $(x_1, p_1) \in \overline{S}$ et $(x_2, p_2) \in S$.

Preuve :

Puisque h^1 et h^2 sont deux fonctions de Bregman, il est évident que h est continûment différentiable sur S, strictement convexe sur $\overline{S}$ et continue sur $\overline{S}$.

De plus, il suit des définitions de h, D_{h^1} et D_{h^2} que pour tout $(x_1, p_1) \in \overline{S}$ et $(x_2, p_2) \in S$,

$$\begin{aligned}
D_h((x_1,p_1),(x_2,p_2)) &= h^1(x_1) + h^2(p_1) - h^1(x_2) - h^2(p_2) \\
&\quad - < \nabla h^1(x_2), x_1 - x_2 > - < \nabla h^2(p_2), p_1 - p_2 > \\
\Leftrightarrow D_h((x_1,p_1),(x_2,p_2)) &= D_{h^1}(x_1,x_2) + D_{h^2}(p_1,p_2)
\end{aligned}$$

La thèse résulte alors des propriétés des fonctions "distance" D_{h^1} et D_{h^2} associées aux fonctions de Bregman h^1 et h^2. ■

Voici à présent quelques exemples de fonction de Bregman et la mesure de "distance" D_h correspondante.

Exemple 2.7 :
La fonction h définie de $\mathbb{R}^n$ dans $\mathbb{R}$ par $h(x) = \frac{1}{2}||x||^2$ où $||.||$ désigne la norme euclidienne sur $\mathbb{R}^n$, est une fonction de Bregman de zone $\mathbb{R}^n$.

Avant de prouver ce résultat, constatons que la fonction D_h est définie de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$ par $D_h(x,y) = \frac{1}{2}||x-y||^2$.
En effet, par définition,

$$\forall\ (x,y) \in \mathbb{R}^n \times \mathbb{R}^n \quad D_h(x,y) = h(x) - h(y) - < \nabla h(y), x - y > \ .$$

Or $\nabla h(y) = y$, d'où $D_h(x,y) = \frac{1}{2}||x||^2 - \frac{1}{2}||y||^2 - < y, x-y >$,
ou, de façon équivalente,

$$D_h(x,y) = \frac{1}{2}||x-y||^2\ . \tag{2.6}$$

Preuve :

Remarquons d'abord que la fonction h est deux fois continûment différentiable et strictement convexe sur $\mathbb{R}^n$. En effet, puisque $\nabla^2 h(x) = I$ pour tout $x \in \mathbb{R}^n$, il suit que l'opérateur $\nabla^2 h$ est défini positif sur $\mathbb{R}^n$ et donc h est strictement convexe sur $\mathbb{R}^n$ (cfr proposition A.I.11).

D'autre part, il est évident que

$$\lim_{||x|| \to \infty} \frac{||x||^2}{||x||} = +\infty .$$

La thèse découle alors du théorème 2.3.

<u>**Exemple 2.8 :**</u>
Soit $\Omega^+ = \{x \in I\!R^n$ tel que $x > 0\}$ et *ent* la fonction entropique définie de $\overline{\Omega}^+$ dans $I\!R$ par $ent\ x = -\sum_{i=1}^n x_i \ln \frac{x_i}{a_i}$ où $a \in \Omega^+$. Notons que par convention : $0 \ln 0 = 0$.
Alors la fonction h définie par $h(x) = -entx$ est une fonction de Bregman de zone Ω^+.

Avant de prouver cette assertion, constatons que la fonction D_h est définie de $\overline{\Omega}^+ \times \Omega^+$ dans $I\!R$ par

$$D_h(x,y) = \sum_{i=1}^n (x_i \ln \frac{x_i}{y_i} - x_i + y_i) . \tag{2.7}$$

En effet, par définition $\forall\ (x,y) \in \overline{\Omega}^+ \times \Omega$

$$D_h(x,y) = h(x) - h(y) - < \nabla h(y), x - y > .$$

Or $\nabla h(y) = (\ln(\frac{y_i}{a_i}) + 1)_{i=1}^n$, d'où

$$D_h(x,y) = \sum_{i=1}^n x_i \ln \frac{x_i}{a_i} - \sum_{i=1}^n y_i \ln \frac{y_i}{a_i} - \sum_{i=1}^n (\ln(\frac{y_i}{a_i}) + 1)(x_i - y_i) .$$

Après simplification, il suit que

$$D_h(x,y) = \sum_{i=1}^n x_i \ln\left(\frac{x_i}{a_i}\right) - \sum_{i=1}^n x_i \ln\left(\frac{y_i}{a_i}\right) - \sum_{i=1}^n x_i + \sum_{i=1}^n y_i ,$$

ou de façon équivalente,

$$D_h(x,y) = \sum_{i=1}^n \left(x_i \ln \frac{x_i}{y_i} - x_i + y_i\right) .$$

<u>**Preuve :**</u>

Nous vérifierons successivement les assertions de la définition d'une fonction de Bregman (cfr définition 2.11). S.p.d.g., nous supposerons que $a_i = 1\ \forall\ i = 1 \cdots n$.

1. La fonction h est continûment différentiable sur Ω^+.

   En effet, $h(x) = \sum_{i=1}^n f(x_j)$ où la fonction $f : I\!R_0^+ \to I\!R$, définie par $f(x) = x \ln x$, est continûment différentiable sur $I\!R_0^+$.

2. La fonction h est continue sur $\overline{\Omega}^+$.

En effet, en vertu de 1), il suffit de prouver que h est continue en $0 \in I\!R^n$. Considérons donc une suite $\{x^k\} \subseteq I\!R^n$ telle que $\lim_{k\to\infty} x^k = 0$.
Par la définition de h, nous obtenons que $\lim_{k\to\infty} h(x^k) = \lim_{k\to\infty} \sum_{i=1}^n x_i^k \ln x_i^k$ d'où $\lim_{k\to\infty} h(x^k) = 0$.
D'autre part, il suit de la convention $0 \ln 0 = 0$ que $h(0) = 0$.
Dès lors, $\lim_{k\to\infty} h(x^k) = h(0)$. La thèse découle alors du critère de réduction aux suites.

3. La fonction h est strictement convexe sur $\overline{\Omega}^+$.

En effet, $h(x) = \sum_{i=1}^n f(x_i)$ où la fonction $f : I\!R^+ \to I\!R$ est définie par $f(x) = x \ln x$ avec la convention $0 \ln 0 = 0$. Montrons que f est strictement convexe sur $I\!R^+$.

- D'une part, nous savons que $\forall\, x \in I\!R_0^+$,
$$\frac{d^2 f(x)}{d^2 x} = \frac{1}{x} > 0\ .$$
Ainsi, il suit de la propriété A.I.10 que f est strictement convexe sur $]I\!R_0^+[$.
- D'autre part, vérifions que $\forall\, \lambda \in]0,1[$ et $\forall\, x > 0$ on a
$$f(\lambda x + (1-\lambda)0) \leq \lambda f(x) + (1-\lambda) f(0)\ .$$
En effet, par la définition de f, cette inégalité se réécrit
$$\lambda x \ln \lambda x \leq \lambda x \ln x\ ,$$
ou encore,
$$\ln \lambda x < \ln x\ .$$
Ce qui est évidemment vérifié car $\lambda x < x$ et la fonction ln est strictement croissante.

4. $\forall\, \alpha \in I\!R$ et $\forall\, y \in \Omega^+$, les ensembles de niveau $L_1(y, \alpha)$ sont bornés.

En effet, supposons *par l'absurde* qu'il existe $y \in \Omega^+$ et $\alpha \in I\!R$ tel que l'ensemble $L_1(y,\alpha)$ contienne une suite $\{x^k\}_{k=0}^\infty$ non bornée.
Dans ce cas, montrons que
$$\lim_{k\to\infty} D_h(x^k, y) = \lim_{k\to\infty} \sum_{i=1}^n (x_i^k \ln x_i^k - x_i^k \ln y_i - x_i^k + y_i) = +\infty\ .$$

- D'une part nous savons que

$$x_i^k \ln x_i^k - x_i^k \ln y_i - x_i^k + y_i \geq 0 \quad \forall\, i = 1 \cdots n \tag{2.8}$$

En effet, la fonction $f : \mathbb{R}^+ \to \mathbb{R}$ définie par $f(x) = x \ln x$ (avec $0 \ln 0 = 0$) est continûment différentiable sur $\mathbb{R}_0^+$, strictement convexe sur $\mathbb{R}^+$ et continue sur $\mathbb{R}^+$. Ainsi, la positivité de la mesure de distance D_f correspondante nous assure que

$$D_f(x_i^k, y_i) = x_i^k \ln x_i^k - x_i^k \ln y_i - x_i^k + y_i \geq 0 \qquad \forall\, i \in 1, \cdots, n\,.$$

- D'autre part, la suite $\{x^k\}_{k=0}^{\infty}$ étant non bornée, il existe un indice $j \in 1, \cdots, n$ tel que la suite des composantes $\{x_j^k\}_{k=0}^{\infty}$ est non bornée. Plus précisément, nous pouvons en extraire une sous-suite, que sans perte de généralité nous notons $\{x_j^k\}_{k=0}^{\infty}$, telle que $\lim_{k\to\infty} x_j^k = +\infty$.
Par conséquent, passant à la limite sur les indices de la sous-suite $\{x_j^k\}$, nous obtenons que

$$\lim_{k\to\infty} (x_j^k \ln x_j^k - x_j^k \ln y_j - x_j^k + y_j) = +\infty \tag{2.9}$$

où y_j désigne la composante de y correspondante à la suite de composantes $\{x_j^k\}$.

- En conclusion, il résulte de (2.8) et (2.9) que

$$\lim_{k\to\infty} D_h(x^k, y) = \lim_{k\to\infty} \sum_{i=1}^{n} (x_i^k \ln x_i^k - x_i^k \ln y_i - x_i^k + y_i) = +\infty\,. \tag{2.10}$$

Or l'hypothèse $\{x^k\} \subseteq L^1(y, \alpha)$, affirme que

$$\lim_{k\to\infty} D_h(x^k, y) \leq \alpha\,.$$

Ce qui contredit (2.10).

Par un raisonnement analogue, nous pouvons montrer que $\forall\, \beta \in \mathbb{R}$ et $\forall\, x \in \overline{\Omega}^+$, les ensembles de niveau $L_2(x, \beta)$ sont bornés.

5. Soit une suite $\{y^k\}_{k=0}^{\infty} \subseteq \Omega^+$ telle que $\lim_{k\to\infty} y^k = y^*$.

Il suit immédiatement de la continuité de la fonction ln que

$$\lim_{k\to\infty} D_h(y^*, y^k) = \sum_{i=0}^{n} \lim_{k\to\infty} (y_i^* \ln \frac{y_i^*}{y_i^k} + y_i^k - y_i^*) = 0\,.$$

6. Soient deux suites $\{x^k\}_{k=0}^{\infty}$ et $\{y^k\}_{k=0}^{\infty}$ contenues dans Ω^+ telles que $\{x^k\}$ est non bornée, $\lim_{k\to\infty} y^k = y^*$ et $\lim_{k\to\infty} D_h(x^k, y^k) = 0$.
Montrons que $\lim_{k\to\infty} x^k = y^*$.

Soit $x^* \in \overline{\Omega}^+$ une valeur d'adhérence de la suite $\{x^k\}$. Notons $\{x^{qk}\}$ la sous-suite extraite de $\{x^k\}$ telle que $\lim_{k\to\infty} x^{qk} = x^*$.

- Il suit que la continuité de la fonction ln que

$$\lim_{k\to\infty} D_h(x^{qk}, y^{qk}) = \lim_{k\to\infty} \sum_{i=1}^{n} (x_i^{qk} \ln \frac{x_i^{qk}}{y_i^{qk}} - x_i^{qk} + y_i^{qk}) = \sum_{i=1}^{n} (x_i^* \ln \frac{x_i^*}{y_i^*} - x_i^* + y_i^*) .$$

- Evaluons à présent le terme général de cette somme. Soit $j \in 1 \cdots n$, un indice arbitraire.
Si $x_j^* = y_j^*$ alors

$$x_j^* \ln \frac{x_j^*}{y_j^*} - x_j^* + y_j^* = 0 .$$

si $x_j^* = 0$ et $y_j^* > 0$ alors

$$x_j^* \ln \frac{x_j^*}{y_j^*} - x_j^* + y_j^* = y_j^* > 0 .$$

si $x_j^* > 0$ et $y_j^* = 0$ alors

$$x_j^* \ln \frac{x_j^*}{y_j^*} - x_j^* + y_j^* = +\infty .$$

- Cependant, par hypothèse, $\lim_{k\to\infty} D_h(x^{qk}, y^{qk}) = \sum_{i=1}^{n} (x_i^* \ln \frac{x_i^*}{y_i^*} - x_i^* + y_i^*) = 0$.
Ainsi, la seule situation autorisée est $x_i^* = y_i^* \ \forall\, i = 1 \cdots n$, c'est-à-dire $x^* = y^*$.
Par conséquent, toute valeur d'adhérence de la suite $\{x^k\}$ est égale à y^*, autrement dit, $\lim_{k\to\infty} x^k = y^*$. Ce qu'il fallait démontrer. ■

Exemple 2.9 :
Soit $\Omega^+ = \{x \in I\!R^n \text{ tel que } x > 0\}$.
La fonction h définie de $\overline{\Omega}^+$ dans $I\!R$ par $h(x) = \sum_{i=1}^{n} (x_i \ln x_i - x_i)$ est une fonction de Bregman de zone Ω^+.
De plus la fonction D_h est définie de $\overline{\Omega}^+ \times \Omega^+$ dans $I\!R$ par

$$D_h(x, y) = \sum_{i=1}^{n} (x_i \ln \frac{x_i}{y_i} - x_i + y_i) . \qquad (2.11)$$

Preuve :

Il est évident que $h(x) = h'(x) + < c, x > +b$ où $c_i = 1 \quad i = 1, \cdots, n \quad , b = 0$ et la fonction h' est définie de $\overline{\Omega}^+$ dans $I\!R$ par $h'(x) = \sum_{i=1}^n \; x_i \ln x_i$.

Or la fonction h' est une fonction de Bregman de zone Ω^+ (cfr exemple 2.8).
Par conséquent, il résulte du théorème 2.4 que h est une fonction de Bregman de zone Ω^+ telle que

$$D_h(x,y) = D_{h'}(x,y) = \sum_{i=1}^{n} (x_i \ln \frac{x_i}{y_i} - x_i + y_i) \; .$$

Ce qu'il fallait démontrer. ∎

Exemple 2.10 :
Soit $S = \{x \in I\!R^n \text{ tel que } x_i > -1 \quad i = 1, \cdots, n\}$.
La fonction h définie de $\overline{S}$ dans $I\!R$ par $h(x) = \sum_{i=1}^n \; (x_i+1)[\ln(x_i+1) - 1]$ est une fonction de Bregman de zone S.

Preuve :
Soit le changement de variable $y = x + 1$. La thèse suit alors de l'exemple 2.9. ∎

2.4.2 Forme de l'itération de l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman

Rappelons que nous désirons résoudre le problème suivant :

(P) trouver $x \in I\!R^n$ tel que $0 \in T(x)$

où T est un opérateur maximal monotone sur $I\!R^n$.

Comme dans le cas classique, le problème (P) revient à trouver un point fixe de l'opérateur $T := (\nabla h + \lambda T)^{-1} \circ \nabla h$ quelle que soit la fonction de Bregman h de zone S et quel que soit $\lambda > 0$.

En effet, si nous acceptons pour un moment que P est univoque, nous avons successivement :

$$\begin{aligned} x = ((\nabla h + \lambda T)^{-1} \circ \nabla h)(x) &\Leftrightarrow \nabla h(x) \in \nabla h(x) + \lambda T(x) \\ &\Leftrightarrow 0 \in \lambda T(x) \\ &\Leftrightarrow 0 \in T(x) \end{aligned}$$

D'où l'idée d'écrire l'itération du point fixe pour résoudre le problème (P). Plus précisément, nous définissons l'algorithme suivant :

<u>Algorithme 2.1 :</u>

Soient $\{\lambda_k\}$ une suite de scalaires strictement positifs et h une fonction de Bregman de zone $S \subseteq I\!R^n$ telle que $D(T) \subseteq S$.
Etant donné $x^0 \in S$, *l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman* génère une suite $\{x^k\}_{k=0}^{\infty}$ telle que

$$x^{k+1} = ((\nabla h + \lambda_k T)^{-1} \circ \nabla h)(x^k) . \tag{2.12}$$

Avant d'étudier la convergence de la suite $\{x^k\}$, donnons quelques propriétés importantes de l'opérateur P, notamment que P est univoque.

<u>Proposition 2.6 :</u>
Soient T un opérateur monotone sur $I\!R^n$ et h une fonction de Bregman de zone $S \subseteq I\!R^n$, telle que $D(T) \subseteq S$. Alors l'opérateur $P := (\nabla h + T)^{-1} \circ \nabla h$ est univoque.

<u>Preuve :</u>

D'une part l'opérateur ∇h est strictement monotone car h est une fonction strictement convexe sur $\bar{S}$ (cfr exemple 2.3).

D'autre part, par hypothèse T est un opérateur monotone, dès lors l'opérateur $\nabla h + T$ est strictement monotone, c'est-à-dire

$$\forall\, x, y \in D(T) \text{ tels que } x \neq y,\ \forall\, y_1 \in (\nabla h + T)(x)\ ,\quad \forall\, y_2 \in (\nabla h + T)(y)$$

$$< y_1 - y_2, x - y > > 0$$

Par conséquent, pour tout $x, y \in D(T)$ tels que $x \neq y$, nous obtenons que

$$(\nabla h + T)(x) \cap (\nabla h + T)(y) = \emptyset\ ,$$

et donc l'opérateur $(\nabla h + T)^{-1}$ est univoque.
La thèse découle alors du fait que ∇h est univoque. ∎

Proposition 2.7 :
Soient $X \subseteq I\!R^n$ un ensemble convexe, fermé, non vide et h une fonction de Bregman de zone S telle que $X \subseteq S$. Considérons un vecteur $y \in I\!R^n \setminus X$.
Alors $Py := ((\nabla h + N_X)^{-1} \circ \nabla h)(y)$ est la D_h-projection de y sur X, où $N_X(y)$ désigne le cône normal à X en y.

Preuve :

1. Montrons d'abord que l'opérateur P est bien défini, c'est-à-dire que N_X est un opérateur monotone.

   Puisque X est un ensemble convexe, fermé, non vide; sa fonction indicatrice est propre, convexe et fermée (cfr proposition A.III.5). Dès lors $\partial\delta_X$ est un opérateur monotone (cfr exemple 2.2). Ainsi, puisque $N_X = \partial\delta_X$, nous déduisons que N_X est un opérateur monotone.

2. Montrons ensuite que $Py = x^*$ où $x^* = \text{argmin}_{\ x\in X}\ \{D_h(x, y)\}$.

   - On a immédiatement que $x^* = \text{argmin}_{\ x\in I\!R^n}\ \{D_h(x, y) + \delta_X(x)\}$. Ainsi, écrivant la condition d'optimalité de ce problème de minimisation, nous obtenons que

     $$0 \in \partial\{D_h(., y) + \delta_X(.)\}(x^*)\ ,$$

     c'est-à-dire

     $$0 \in \partial\{h(.) - h(y) - < \nabla h(y), . - y > + \delta_X(.)\}(x^*)\ . \tag{2.13}$$

- Or $ri(X) \cap ri(\text{dom } h) \neq \emptyset$ (où l'abréviation ri désigne l'intérieur relatif (cfr définition A.II.1)). En effet, d'une part $ri(X) \subseteq S$ car $X \subseteq S$ et S est un ouvert. D'autre part $ri(X) \neq \emptyset$ car X est convexe (cfr proposition A.II.4).

- Par conséquent, il suit des règles de calcul du sous-différentiel (cfr proposition A.IV.12) que la relation (2.13) est équivalente à

$$0 \in \nabla h(x^*) - \nabla h(y) + \partial \delta_X(x^*)$$

c'est-à-dire,

$$\nabla h(y) \in (\nabla h + N_X)(x^*) \,,$$

ou encore,

$$x^* \in ((\nabla h + N_X)^{-1} \circ \nabla h)(y) \,. \tag{2.14}$$

- De plus, puisque N_X est un opérateur monotone, il suit de la proposition 2.6, que l'opérateur $((\nabla h + N_X)^{-1} \circ \nabla h)$ est univoque. Ainsi nous pouvons écrire

$$x^* = ((\nabla h + N_X)^{-1} \circ \nabla h)\ (y)$$

c'est-à-dire la thèse. ∎

Considérons à présent la propriété de "non expansivité" de l'opérateur P.

Proposition 2.8 :
Soient T un opérateur monotone sur $I\!R^n$, h une fonction de Bregman de zone $S \subseteq I\!R^n$ telle que $D(T) \subseteq S$ et P l'opérateur défini par $P := (\nabla h + T)^{-1} \circ \nabla h$. Considérons un vecteur $y \in I\!R^n$ tel que Py existe.
Si $z \in \overline{S}$ est un zéro de T alors $D_h(z, Py) \leq D_h(z, y) - D_h(Py, y)$.

Preuve :

1. Montrons d'abord que la thèse est équivalente à la relation suivante $< \nabla h(y) - \nabla h(Py), Py - z > \geq 0$.
   Pour cela, il suffit de remarquer l'équivalence des inégalités suivantes :

$$\begin{aligned}
D_h(z, Py) &\leq D_h(z, y) - D_h(Py, y) \\
\Leftrightarrow - < \nabla h(Py), z - Py > &\leq - < \nabla h(y), z - y > + < \nabla h(y), Py - y > \\
\Leftrightarrow < \nabla h(Py), Py - z > &\leq < \nabla h(y), y - z > + < \nabla h(y), Py - y > \\
\Leftrightarrow < \nabla h(Py), Py - z > &\leq < \nabla h(y), Py - z > \\
\Leftrightarrow < \nabla h(y) - \nabla h(Py), Py - z > &\geq 0 \,.
\end{aligned}$$

2. Montrons ensuite que $< \nabla h(y) - \nabla h(Py), Py - z > \geq 0$.
Par définition de P on a que

$$\nabla h(y) \in (\nabla h + T)(Py) ,$$

c'est-à-dire

$$\nabla h(y) - \nabla h(Py) \in T(Py) .$$

Comme, d'autre part $0 \in Tz$, il suit de la monotonie de T que

$$< \nabla h(y) - \nabla h(Py) - 0, Py - z > \geq 0$$

Ce qu'il fallait démontrer. ■

2.4.3 Conditions d'existence des itérés de l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman

Dans ce paragraphe, nous désirons analyser les conditions d'existence de la suite $\{x^k\}_{k=0}^{\infty}$ générée par l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman, c'est-à-dire

$$x^{k+1} = ((\nabla h + \lambda_k T)^{-1} \circ \nabla h)(x^k)$$

où $\{\lambda_k\}$ est une suite de scalaires strictement positifs et h une fonction de Bregman de zone $S \subseteq I\!R^n$ telle $D(T) \subseteq S$.

Enonçons donc le théorème suivant :

Théorème 2.6 :
Soient T un opérateur maximal monotone sur $I\!R^n$ et h une fonction de Bregman de zone S telle que $D(T) \subseteq S$.
Si l'une des deux conditions suivantes est vérifiée :

H.2.1. l'image du gradient de h est égal à $I\!R^n$.

H.2.2. l'image du gradient h est un ouvert et 0 appartient à l'image de T.

Alors la suite $\{x^k\}$ générée par l'itération

$$x^{k+1} = ((\nabla h + T)^{-1} \circ \nabla h)(x^k)$$

est bien définie pour tout $x^0 \in S$.

Dans le but de prouver ce résultat, rappelons tout d'abord une propriété de l'image de la somme de deux opérateurs monotones.

Définition 2.12 :
Soit T un opérateur monotone. On dit que T vérifie la *L-propriété* si

$$\forall\, u \in D(T),\ \forall\, v \in R(T)\ \inf\ \{< x-u, y-v > \text{ tel que } x \in D(T), y \in Tx\} > -\infty$$

Proposition 2.9 :
Si T un opérateur trimonotone, alors T vérifie la L-propriété.

Preuve : cfr [3], p. 392.

Proposition 2.10 :
Soient A et B deux opérateurs monotones tels que

1. $D(A) \subseteq D(B)$
2. $A+B$ est maximal monotone
3. B vérifie la L- propriété

alors

$$\text{int}\ (R(A+B)) = \text{int}\ (R(A)+R(B))$$

et

$$R(A+B) = R(A)+R(B)\ .$$

Preuve : cfr [3], théorème 17, p. 393.

Preuve du théorème 2.6 :

1. Prouvons d'abord que int $(R(T+\nabla h)) =$ int $(R(T)+R(\nabla h))$.

   Remarquons dès à présent que la fonction h étant différentiable et convexe sur S (def. 2.11, (i) et (ii)), son gradient est un opérateur maximal monotone et trimonotone de domaine S (exemple 2.5, exemple 2.6).
   Nous déduisons alors les propriétés suivantes :

   - D'une part, l'opérateur $T+\nabla h$ est maximal monotone.

     En effet, ∇h et T sont deux opérateurs maximaux monotones. (2.15)

De plus, par hypothèse $ri(D(T)) \subseteq D(T) \subseteq S$ et $ri(D(\nabla h)) = S$,

$$\text{d'où } ri(D(T)) \subseteq ri(D(\nabla h)) \; . \tag{2.16}$$

Or $D(T)$ est convexe puisque T est maximal monotone (proposition 2.3), ainsi, il suit d'une propriété de l'intérieur relatif (proposition A.II.4) que

$$ri(D(T)) \neq \emptyset \; . \tag{2.17}$$

Dès lors, il suit des relations (2.16) et (2.17) que

$$ri(D(\nabla h)) \cap ri(D(T)) \neq \emptyset \; . \tag{2.18}$$

Par conséquent, nous déduisons des relations (2.15), (2.18) et du théorème 2.1 que $\nabla h + T$ est un opérateur maximal monotone.

- D'autre part, la proposition 2.9 nous assure que l'opérateur trimonotone ∇h vérifie la $L-$propriété.
- Enfin, puisque par hypothèse que $D(T) \subseteq S$, nous obtenons que $D(T) \subseteq D(\nabla h)$.

La thèse suit alors de la proposition 2.10 où nous identifions $A := T$ et $B := \nabla h$.

2. Vérifions ensuite que l'opérateur $(\nabla h + T)^{-1} \circ \nabla h$ est bien défini sur $D(\nabla h)$, c'est-à-dire que

$$R(\nabla h) \subseteq D(\nabla h + T)^{-1} = R(\nabla h + T) \; .$$

Nous savons déjà par la première partie de la preuve que

$$\text{int } (R(T + \nabla h)) = \text{ int } (R(T) + R(\nabla h)). \tag{2.19}$$

Dès lors, si $R(\nabla h) = I\!R^n$ nous obtenons que int $(R(T) + R(\nabla h)) = I\!R^n$ et donc $R(\nabla h) = I\!R^n = R(T + \nabla h)$.

Par ailleurs si $R(\nabla h)$ est un ensemble ouvert et $0 \in R(T)$, alors

$$\text{int } (\{0\} + R(\nabla h)) \subseteq \text{ int } (R(T) + R(\nabla h)) \; ,$$

autrement dit

$$\text{int } (R(\nabla h)) \subseteq \text{ int } (R(T) + R(\nabla h)) \; .$$

Ainsi nous déduisons de la relation (2.19) que

$$\text{int } (R(\nabla h)) \subseteq \text{ int } R(T + \nabla h) \subseteq R(T + \nabla h) \; ,$$

Or $R(\nabla h)$ est un ensemble ouvert, par conséquent nous obtenons que $R(\nabla h) \subseteq R(T + \nabla h)$.

3. Finalement, il suit des définitions de l'inverse d'un opérateur, et de la somme de deux opérateurs que
$$R(\nabla h + T)^{-1} = D(\nabla h + T) \subseteq D(\nabla h)\ .$$
De plus, en vertu de la deuxième partie de la preuve, l'opérateur $(\nabla h + T)^{-1} \circ \nabla h$ est bien défini sur $D(\nabla h)$.
Par induction, nous pouvons donc affirmer que pour tout $x_0 \in D(\nabla h) = S$ et $\forall\ k \geq 0, x^{k+1} = ((\nabla h + T)^{-1} \circ \nabla h)(x^k)$ existe. Ce qu'il fallait démontrer. ∎

Remarque 2.1 :

Dans la suite de ce mémoire, nous supposerons que l'une des deux hypothèses **(H.2.1.)** ou **(H.2.2.)** est vérifiée.
Appliquant alors le théorème 2.6 à l'opérateur maximal monotone $\lambda_k T$ où $\lambda_k > 0$, nous déduisons *l'existence de la suite définie par l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman.*

Vérifions à présent que les quatre exemples de fonction de Bregman mentionnés dans le paragraphe précédent vérifient l'hypothèse **(H.2.1)** du théorème 2.6, c'est-à-dire $R(\nabla h) = \mathbb{R}^n$.

Exemple 2.11 :
Si la fonction h est définie de $\mathbb{R}^n$ dans $\mathbb{R}$ par $h(x) = \frac{1}{2}||x||^2$, alors $\nabla h(x) = x$. D'où $R(\nabla h) = \mathbb{R}^n$.

Exemple 2.12 :
Si la fonction h est définie de $\overline{\Omega}^+$ dans $\mathbb{R}$ par $h(x) = \sum_{i=1}^{n} \ x_i \ln x_i$ avec la convention $0 \ln 0 = 0$, alors $\nabla h(x) = (\ln x_i + 1)_{i=1}^n$.
Il est donc immédiat que $R(\nabla h) \subseteq \mathbb{R}^n$.
D'autre part $\mathbb{R}^n \subseteq R(\nabla h)$. En effet, pour tout $y \in \mathbb{R}^n$, il suffit de considérer $x \in \Omega^+$ tel que $x_i = e^{y_i - 1} \quad i = 1 \cdots n$, pour obtenir $y_i = \ln x_i + 1 \quad i = 1 \cdots n$.

Exemple 2.13 :
Si la fonction h est définie de $\overline{\Omega}^+$ dans $\mathbb{R}$ par $h(x) = \sum_{i=1}^{n} (x_i \ln x_i - x_i)$, alors $\nabla h(x) = (\ln x_i)_{i=1}^n$.

Par un raisonnement analogue à celui de l'exemple 2.12, il suit que $R(\nabla h) = I\!R^n$.

Exemple 2.14 :
Soit $S = \{x \in I\!R^n$ tel que $x_i > -1$ pour tout $i = 1, \cdots, n\}$.
Si la fonction h est définie de $\overline{S}$ dans $I\!R$ par $h(x) = \sum_{i=1}^{n} (x_i + 1)[\ln(x_i + 1) - 1]$,
alors $\nabla h(x) = (\ln(x_i + 1))_{i=1}^{n}$
Par un raisonnement analogue à celui de l'exemple 2.12, il suit que $R(\nabla h) = I\!R^n$.

2.4.4 Convergence de l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman

Remarquons tout d'abord que par le changement de variable non linéaire $y^k = \nabla h(x^k)$, l'itération $x^{k+1} = ((\nabla h + \lambda_k T)^{-1} \circ \nabla h)(x^k)$ se réécrit $y^{k+1} = (I + \lambda_k (T \circ (\nabla h)^{-1}))^{-1}(y^k)$. Laquelle est bien sûr l'itération de l'algorithme du point proximal linéaire appliqué à l'opérateur $T \circ (\nabla h)^{-1}$ (cfr paragraphe 2.3). Cependant, les compositions d'opérateurs monotones n'étant pas nécessairement monotones, nous ne pouvons pas appliquer la théorie de convergence établie dans le cadre de l'algorithme du point proximal linéaire.

Enonçons à présent le théorème de convergence de la suite $\{x^k\}$ définie par l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman (algorithme 2.1) vers une solution du problème (P).

Théorème 2.7 : *Convergence vers une solution du problème P*
Soient T un opérateur maximal monotone sur $I\!R^n$, h une fonction de Bregman de zone S telle que $\overline{D(T)} \subseteq S$, et $\{\lambda_k\}_{k=0}^{\infty}$ une suite de scalaires strictement positifs telle que $\inf_{k \geq 0} \{\lambda_k\} > 0$.
Considérons également la suite $\{x^k\}_{k=0}^{\infty} \subseteq I\!R^n$ définie par l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman, c'est-à-dire

$$x^{k+1} = ((\nabla h + \lambda_k T)^{-1} \circ \nabla h)(x^k) \quad \forall\, k \geq 0\,.$$

. Dès lors la suite $\{x^k\}$ converge vers un zéro de l'opérateur T s'il en existe un. Dans le cas contraire, la suite $\{x^k\}$ est non bornée.

Dans le but de prouver ce théorème, rappelons tout d'abord une propriété des suites convergentes.

Lemme 2.1 :

Soit $\{x^k\}_{k=0}^{\infty}$ une suite décroissante de $I\!R^n$.
Si $\{x^{k_j}\}_{j=0}^{\infty}$ est une sous-suite extraite de $\{x^k\}$ telle que $\lim_{j\to\infty} x^{k_j} = 0$, alors $\lim_{k\to\infty} x^k = 0$.

Preuve du lemme 2.1 :

- Montrons d'abord que la suite $\{x^k\}$ converge.
  En effet, puisque la sous-suite $\{x^{k_j}\}$ décroit et converge vers 0, nous obtenons que

$$x^{k_j} \geq 0 \qquad \forall\, j \in I\!N \tag{2.20}$$

  Par ailleurs, il suit de la croissance des indices d'une sous-suite que $k_j \geq k \; \forall\, k, j \in I\!N$.
  Ainsi, nous déduisons de la décroissance de la suite $\{x^k\}$ que

$$x^{k_j} \leq x^k \qquad \forall\, k, j \in I\!N \tag{2.21}$$

  Regroupant alors les inégalités (2.20) et (2.21), nous obtenons que

$$0 \leq x^{k_j} \leq x^k \quad \forall\, k, j \in I\!N\,.$$

  Par conséquent, la suite $\{x^k\}$ est décroissante et minorée par 0. Il s'agit donc d'une suite convergente.

- Finalement, puisque toutes les sous-suites extraites d'une suite convergente convergent vers la même limite, nous concluons que $\lim_{k\to\infty} x^k = 0$. Ce qu'il fallait démontrer.

Prouvons à présent le théorème de convergence associé au problème (P).

Preuve du théorème 2.7 :

Partie 1. Supposons que l'opérateur T admette un ensemble non vide de zéros, c'est-à-dire qu'il existe $z \in D(T)$ tel que $T(z) = 0$.

Nous désirons alors prouver la convergence de la suite $\{x^k\}$ vers un zéro de T.
Dans ce but, nous vérifierons d'abord que la suite $\{x^k\}$ est bornée (Pas 1). Ensuite, nous prouverons que toutes les valeurs d'adhérence de cette suite sont des zéros de l'opérateur T (Pas 2). Enfin, nous montrerons que la suite $\{x^k\}$ admet une seule valeur d'adhérence appelée valeur limite de la suite (Pas 3).

Pas 1 : [La suite $\{x^k\}$ est bornée]

D'une part, nous savons que l'opérateur $\lambda_k T$ est monotone pour tout $\lambda_k > 0$.
D'autre part, $z \in S$ est un zéro de T.
Ainsi, il résulte de la proposition 2.8 que :

$$D_h(z, x^{k+1}) \leq D_h(z, x^k) - D_h(x^{k+1}, x^k) \ \forall\ k \in \mathbb{N}\,. \tag{2.22}$$

D'où, par la positivité de la fonction D_h, $D_h(z, x^{k+1}) \leq D_h(z, x^k) \quad \forall\ k \in \mathbb{N}$.
Par conséquent, la suite $\{x^k\}$ est contenue dans l'ensemble de niveau

$$L_2(z, D_h(z, x^0)) := \{y \in S \text{ tq } D_h(z, y) \leq D_h(z, x^0)\}\,.$$

Or cet ensemble est borné car h est une fonction de Bregman (définition 2.11, (iv)). Ainsi, nous concluons que la suite $\{x^k\}$ est bornée.

Pas 2 : [Toute valeur d'adhérence de la suite $\{x^k\}$ est un zéro de l'opérateur T]

La suite $\{x^k\}$ étant bornée, nous pouvons en extraire une sous-suite, que nous noterons $\{x^{k_j}\}$, telle que $x^* = \lim_{j\to\infty} x^{k_j}$.

1. Constatons tout d'abord que

$$\lim_{j\to\infty} x^{k_j+1} = x^*\,. \tag{2.23}$$

- En effet, sommant l'inégalité (2.22) répétée pour k variant entre 0 et m (m étant un entier quelconque), nous obtenons que

$$\sum_{k=0}^{m} D_h(x^{k+1}, x^k) \leq D_h(z, x^0)\,.$$

Par conséquent, la série à termes positifs $\sum_{k=0}^{\infty} D_h(x^{k+1}, x^k)$ admet une somme finie, et donc

$$\lim_{k\to\infty} D_h(x^{k+1}, x^k) = 0\,.$$

ou encore

$$\lim_{j\to\infty} D_h(x^{k_j+1}, x^{k_j}) = 0\,.$$

- De plus, la suite $\{x^k\}$ étant bornée, il en est de même pour la sous-suite $\{x^{k_j}\}$.

- Ainsi, il suit de la définition h comme fonction de Bregman (définition 2.12 (vi)) que

$$\lim_{j\to\infty} x^{k_j+1} = x^*$$

2. Réécrivons ensuite la relation d'itération de la suite $\{x^k\}$. Nous avons successivement que

$$x^{k+1} = (\nabla h + \lambda_k T)^{-1}(\nabla h(x^k)) \quad \forall\, k \in \mathbb{N}$$

c'est-à-dire

$$\nabla h(x^k) \in (\nabla h + \lambda_k T)(x^{k+1}) \quad \forall\, k \in \mathbb{N}\,,$$

ou encore

$$\frac{1}{\lambda_k}(\nabla h(x^k) - \nabla h(x^{k+1})) \in T(x^{k+1}) \quad \forall\, k \in \mathbb{N}\,.$$

Appliquant cette relation à la sous-suite $\{x^{k_j}\}$, nous obtenons que

$$\frac{1}{\lambda_{k_j}}(\nabla h(x^{k_j}) - \nabla h(x^{k_j+1})) \in T(x^{k_j+1}) \quad \forall\, k_j \in \mathbb{N}\,. \tag{2.24}$$

3. Montrons maintenant que

$$\lim_{j\to\infty} \frac{1}{\lambda_{k_j}}\,(\nabla h(x^{k_j}) - \nabla h(x^{k_j+1})) = 0\,. \tag{2.25}$$

D'une part, il suit de l'hypothèse $\overline{D(T)} \subseteq S$ que $x^* \in S$.
D'autre part, rappelons que par définition $\lim_{j\to\infty}\; x^{k_j} = x^*$ et que par la relation (2.23) $\lim_{j\to\infty}\; x^{k_j+1} = x^*$.
Dès lors, puisque la fonction de Bregman h est continûment différentiable sur S, nous obtenons que

$$\lim_{j\to\infty}\; \nabla h(x^{k_j}) = \nabla h(x^*) \text{ et } \lim_{j\to\infty}\; \nabla h(x^{k_j+1}) = \nabla h(x^*)\,,$$

d'où

$$\lim_{j\to\infty}(\nabla h(x^{k_j}) - \nabla h(x^{k_j+1})) = 0$$

Or par hypothèse $\inf_{k\geq 0}\{\lambda_k\} > 0$, dès lors,

$$\lim_{j\to\infty} \frac{1}{\lambda_{k_j}}(\nabla h(x^{k_j}) - \nabla h(x^{k_j+1})) = 0$$

4. Finalement, en vertu du caractère fermé des opérateurs maximaux monotones (cfr proposition 2.4) et des relations (2.23), (2.24) et (2.25), pour rappel

$$\begin{aligned}
&\lim_{j\to\infty} x^{k_j+1} = x^* \\
&\frac{1}{\lambda_{k_j}}(\nabla h(x^{k_j}) - \nabla h(x^{k_j+1})) \in T(x^{k_j+1}) \quad \forall\, k_j \in I\!N \\
&\lim_{j\to\infty} \frac{1}{\lambda_{k_j}}(\nabla h(x^{k_j}) - \nabla h(x^{k_j+1})) = 0 \; ,
\end{aligned}$$

nous concluons que $0 \in T(x^*)$.
Autrement dit, toute valeur d'adhérence de la suite $\{x^k\}$ est un zéro de l'opérateur T.

<u>Pas 3 :</u> *[La suite $\{x^k\}$ admet une seule valeur d'adhérence]*

Soit $\tilde{x} \in \overline{S}$ tel que $\tilde{x} \neq x^*$ une deuxième valeur d'adhérence de la suite $\{x^k\}$ et soit $\{x^{k'_j}\}$ la sous-suite convergente correspondante, c'est-à-dire

$$\lim_{j\to\infty} x^{k'_j} = \tilde{x} \; . \tag{2.26}$$

Notre but sera de prouver que $\tilde{x} = x^*$.

1. Remarquons tout d'abord que

$$\lim_{k\to\infty} D_h(x^*, x^k) = 0 \; . \tag{2.27}$$

En effet, par l'inégalité (2.22) appliquée à $z = x^* \in T^{-1}(0)$ nous obtenons que

$$\forall\, k \in I\!N \quad D_h(x^*, x^{k+1}) \leq D_h(x^*, x^k) - D_h(x^{k+1}, x^k) \; .$$

Ainsi, il suit de la positivité de la fonction D_h que la suite $\{D_h(x^*, x^k)\}_{k=0}^{\infty}$ est décroissante.

D'autre part, la sous-suite $\{x^{k_j}\} \subseteq D(T) \subseteq S$ est telle que $\lim_{j\to\infty} x^{k_j} = x^*$.

Par conséquent, nous déduisons de la définition de h comme fonction de Bregman (définition 2.11 point v) que

$$\lim_{j\to\infty} D_h(x^*, x^{k_j}) = 0$$

Le résultat annoncé découle alors du lemme 2.1, où nous identifions x^k à $D_h(x^*, x^k)$ et x^{k_j} à $D_h(x^*, x^{k_j})$.

2. Définissons ensuite $w^k = x^* \ \forall\ k \in I\!N$ et montrons que $\lim_{k\to\infty} \ w^{k_j} = x^*$.

D'une part, puisque la sous-suite $\{D_h(w^{k'_j}, x^{k'_j})\}$ est extraite de la suite $(D_h(x^*, x^k)\}$ et que $\lim_{k\to\infty} \ D_h(x^*, x^k) = 0$, nous obtenons que

$$\lim_{j\to\infty} \ D_h(w^{k'_j}, x^{k'_j}) = 0 \tag{2.28}$$

D'autre part, en vertu du pas 1, la suite $\{x^k\}$ est bornée

$$\text{d'où } \{w^{k'_j}\} \subseteq \{x^k\} \text{ est également une suite bornée.} \tag{2.29}$$

Ainsi, en regroupant les relations (2.26), (2.28), (2.29), pour rappel :

$$\begin{aligned}&\lim_{j\to\infty} \ x^{k'_j} = \tilde{x}\\ &\lim_{j\to\infty} \ D_h(w^{k'_j}, x^{k'_j}) = 0\\ &\text{et la suite } \{w^{k'_j}\} \text{ est bornée}\end{aligned}$$

et appliquant le point (vi) de la définition de h comme fonction de Bregman (définition 2.11), nous obtenons que

$$\lim_{j\to\infty} \ w^{k'_j} = \tilde{x} \ . \tag{2.30}$$

3. Enfin, puisque par définition $w^k = x^* \ \forall\ k \in I\!N$, nous obtenons que $\tilde{x} = x^*$.
   Ce qui nous permet de conclure que la suite $\{x^k\}$ admet une seule valeur d'adhérence x^* qui est un zéro de l'opérateur T.

Partie 2. Supposons maintenant que l'opérateur T ne possède aucun zéro, c'est-à-dire $T^{-1}(0) = \emptyset$.

Nous désirons alors prouver que la suite $\{x^k\}$ est non bornée.

Supposons *par l'absurde* qu'il existe $\delta \in I\!R$ tel que $||x^k|| \leq \delta \quad \forall\ k \in I\!N$.
Soit l'ensemble $C = \{x \in D(T) \text{ tel que } ||x|| \leq \delta\} \subseteq D(T)$ et $p = \sup_{x \in C} \{||\nabla h(x)||\}$.
Il est tout d'abord évident que p est fini. En effet, d'une part C est un ensemble compact inclus dans S car $D(T) \subseteq S$. D'autre part, le gradient de h est une fonction continue sur S(définition 2.11 (i)). Par conséquent, l'opérateur ∇h atteint sa borne supérieure en un

point de $\mathbb{C}$.

Définissons également r :

$$r = \max\{\delta, p\} + 1 \ ,$$

et remarquons que r est fini et $r > \delta$.

Considérons à présent l'opérateur T' défini par

$$T'(x) = \begin{cases} T(x) & \text{si } ||x|| < r \\ \{y + ax \text{ tel que } y \in T(x) \text{ et } a \geq 0\} & \text{si } ||x|| = r \\ \emptyset & \text{si } ||x|| > r \end{cases}$$

et l'ensemble $L = \{x \in \mathbb{R}^n \text{ tel que } ||x|| \leq r\} = D(T')$.

1. Montrons que l'opérateur maximal monotone T' admet au moins un zéro.

   - Dans ce but, remarquons d'abord que $T'(x) = T(x) + \partial g(x) \quad \forall\, x \in \mathbb{R}^n$ où $g(x) = \delta_L(x)$. En effet par définition, nous avons que

   $$\partial g(x) = \begin{cases} N_L(x) = \{y \in \mathbb{R}^n \text{ tq } < y, w - x > \leq 0 \ \forall\, w \in L\} & \text{si } x \in L \\ \emptyset & \text{sinon} \end{cases}$$

   ou plus précisément,

   $$\partial g(x) = \begin{cases} 0 & \text{si } ||x|| < r \\ ax \text{ où } a \geq 0 & \text{si } ||x|| = r \\ \emptyset & \text{si } ||x|| > r \end{cases}$$

   - Montrons ensuite que l'opérateur T' est maximal monotone.
   D'une part, la fonction g est propre et s.c.i. car elle est la fonction indicatrice d'un ensemble convexe fermé L (cfr proposition A.III.5). Son sous-différentiel, ie ∂g, est donc un opérateur maximal monotone (exemple 2.5). Notons également que $D(\partial g) = L$.

   D'autre part, par hypothèse, nous savons que

   $$||x^k|| \leq \delta < r \quad \forall\, k \in \mathbb{N}$$

d'où,

$$x^k \in D(T) \cap \text{ int } D(\partial g) \quad \forall\, k \in I\!N$$

et donc

$$D(T) \cap \text{ int } D(\partial g) \neq \emptyset\,.$$

Finalement, nous déduisons du théorème 2.1 que $T' = T + \partial g$ est un opérateur maximal monotone.

- Enfin, remarquons que T' possède au moins un zéro, c'est-à-dire $(T')^{-1}(0) \neq \emptyset$, car $D(T') = L$ est un ensemble borné [cfr (20), proposition 2, p. 882].

2. Par définition, nous savons que $r > \delta$ d'où $||x^k|| \leq \delta < r \quad \forall\, k \in I\!N$ et donc $T(x^k) = T'(x^k) \quad \forall\, k \in I\!N$.
Ainsi, il suit que

$$(\nabla h + \lambda_k T')^{-1}(\nabla h(x^k)) = (\nabla h + \lambda_k T)^{-1}(\nabla h(x^k)) = x^{k+1} \quad \forall\, k \in I\!N\,.$$

Appliquant alors la partie 1) de la preuve à l'opérateur T', nous obtenons que la suite $\{x^k\}$ converge vers un zéro de l'opérateur T' : soit x^*.

Cependant, rappelons-nous que par hypothèse de l'absurde,

$$||x^k|| \leq \delta < r \quad \forall\, k \in I\!N\,,$$

Dès lors, en passant à la limite, il suit que

$$||x^*|| \leq \delta < r$$

et donc $T(x^*) = T'(x^*)$.

Finalement, nous obtenons que $0 \in T'(x^*) = T(x^*)$ c'est-à-dire que x^* *est un zéro de l'opérateur* T *ce qui contredit l'hypothèse* $T^{-1}(0) = \emptyset$.
L'hypothèse de l'absurde étant fausse, nous concluons que la suite $\{x^k\}$ est bornée.
Ce qu'il fallait démontrer. ■

Remarque 2.2 :

La seule condition restrictive du théorème 2.7 est $\overline{D(T)} \subseteq S$.
Dans beaucoup d'applications, comme par exemple la méthode exponentielle des multiplicateurs (cfr paragraphe 4.4.6), il serait préférable de remplacer cette condition par une hypothèse plus faible.

2.4.5 Cas particulier : $h(x) = \frac{1}{2}||x||^2$

Dans ce paragraphe, nous nous intéressons à la forme de l'itération de l'algorithme du point proximal non linéaire utilisant la fonction de Bregman h est définie de $I\!R^n$ dans $I\!R$ par $h(x) = \frac{1}{2}||x||^2$.

Soit $x^0 \in I\!R^n$. Puisque $\nabla h(x) = x \quad \forall\, x \in I\!R^n$, l'algorithme 2.1 génère une suite $\{x^k\}$ telle que

$$x^{k+1} = (I + \lambda_k T)^{-1}(x^k)$$

où $\{\lambda_k\}$ est une suite de scalaires strictement positifs.

En outre, nous savons par l'exemple 2.11 que $R(\nabla h) = I\!R^n$. Cette suite est donc bien définie quel que soit le point de départ $x^0 \in I\!R^n$ (cfr théorème 2.6).

Nous retrouvons ainsi l'algorithme du point proximal linéaire de Rockafellar rappelé dans le paragraphe 2.3.

Chapitre 3

Algorithme de minimisation proximale utilisant des fonctions de Bregman

Dans ce chapitre, nous désirons appliquer l'algorithme du point proximal non linéaire analysé dans le paragraphe 2.4 au problème convexe (P_1) :

$$(\mathrm{P}_1) \qquad \begin{cases} \text{Minimiser} & f(x) \\ \text{s.c.} & x \in X \end{cases}$$

où f est une fonction propre, convexe et s.c.i., définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$ et X est un ensemble convexe fermé, non vide de $I\!R^n$ tel que dom $f \cap X \neq \emptyset$.

Enonçons tout d'abord deux propositions relatives aux conditions d'existence et d'unicité des solutions du problème (P_1).

3.1 Conditions d'existence et d'unicité des solutions du problème de minimisation convexe (P_1)

Proposition 3.1 :
Si l'ensemble X sur lequel nous minimisons est borné alors il existe au moins une solution au problème (P_1).

Preuve : cfr [22], propriété 2.8, p. 35.

Proposition 3.2 :
Si la fonction f est fortement convexe, alors le problème (P_1) admet une solution unique.

Preuve : cfr [22], propriété 2.9, p. 36.

3.2 Algorithme de minimisation proximale utilisant des fonctions de Bregman

Dans le paragraphe 2.1, nous avons réécrit le problème (P_1) sous la forme de la recherche d'un zéro d'un opérateur maximal monotone. Nous avons obtenu le problème :

$$(\mathrm{P}_4) \qquad \text{Trouver } x \in I\!R^n \text{ tel que } 0 \in \partial f_X(x)$$

où $\partial f_X(x)$ désigne le sous-différentiel de la fonction convexe $f_X := f + \delta_X$ au point x.

Appliquons à présent l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman (algorithme 2.1) au problème (P_4). Nous identifions donc les opérateurs maximaux monotones T et ∂f_X.

Etant donné $x^0 \in S$, nous engendrons ainsi une suite $\{x^k\}_{k=0}^{\infty}$ telle que

$$x^{k+1} = ((\nabla h + \lambda_k \partial f_X)^{-1} \circ \nabla h)(x^k) \tag{3.1}$$

où $\{\lambda_k\}$ est une suite de scalaires strictement positifs et h une fonction de Bregman de zone S telle que $D(\partial f_X) \subseteq S$.

Comme mentionné dans le paragraphe 2.4.3, nous supposerons que l'une des deux conditions suivantes est vérifiée afin d'assurer l'existence de la suite $\{x^k\}$

H.3.1. $R(\nabla h) = I\!R^n$. (3.2)

H.3.2. $R(\nabla h)$ est un ouvert et $0 \in R(\partial f_X)$, ou plus précisément,

$R(\nabla h)$ est un ouvert et la fonction f atteint son minimum sur X . (3.3)

Nous désirons à présent réécrire l'itération (3.1) en terme des fonctions f et D_h. Dans ce but, énonçons le théorème suivant :

<u>**Théorème 3.1 :**</u>
Soient $\{\lambda_k\}$ une suite de scalaires strictement positifs et h une fonction de Bregman de zone S telle que $D(\partial f_X) \subseteq S$. Alors,

$$x^{k+1} = \text{argmin}_{\,x\in X} \; \{f(x) + \frac{1}{\lambda_k} D_h(x, x^k)\} \Leftrightarrow x^{k+1} = ((\nabla h + \lambda_k \partial f_X)^{-1} \circ \nabla h)(x^k) \; .$$

<u>**Preuve :**</u>
Comme

$$x^{k+1} = \text{argmin}_{\,x\in X} \; \{f(x) + \frac{1}{\lambda_k} D_h(x, x^k)\} \; ,$$

il suit de la définition de la fonction D_h que

$$x^{k+1} = \text{argmin}_{\,x\in I\!R^n} \; \{f_X(x) + \frac{1}{\lambda_k}(h(x) - < \nabla h(x^k), x >)\} \tag{3.4}$$

Ainsi, la condition d'optimalité d'un problème de minimisation convexe (cfr proposition A.VII.1), nous assure que la relation (3.4) est équivalente à

$$0 \in \partial\{f_X(.) + \frac{1}{\lambda_k}(h(.) - < \nabla h(x^k), . >)\}(x^{k+1}) \; . \tag{3.5}$$

Or $ri(\text{dom} \;\; f_X) \cap ri(\text{dom} \;\; h) \neq \emptyset$. En effet, d'une part $ri(\text{dom} \;\; h) \neq \emptyset$ car dom h est un ensemble convexe. D'autre part, il suit de l'hypothèse $D(\partial f_X) \subseteq S =$dom h et de l'inclusion $ri(\text{dom} \;\; f_X) \subseteq D(\partial f_X)$ (cfr proposition A.IV.9) que $ri(\text{dom} \;\; f_X) \subseteq S$.

Par conséquent, nous déduisons de la propriété de sommation du sous-différentiel (cfr proposition A.IV.12) que la relation (3.5) est équivalente à

$$0 \in \partial f_X(x^{k+1}) + \partial(\frac{1}{\lambda_k} h(.) - \frac{1}{\lambda_k} < \nabla h(x^k), . >)(x^{k+1}) \; ,$$

c'est-à-dire par la différentiabilité de h,

$$0 \in \partial f_X(x^{k+1}) + \frac{1}{\lambda_k} \nabla h(x^{k+1}) - \frac{1}{\lambda_k} \nabla h(x^k) \; ,$$

ou encore

$$\nabla h(x^k) \in (\nabla h + \lambda_k \partial f_X)(x^{k+1}) ,$$

et donc

$$x^{k+1} \in ((\nabla h + \lambda_k \partial f_X)^{-1} \circ \nabla h)(x^k) . \tag{3.6}$$

De plus, puisque l'opérateur ∂f_X est monotone, il suit de la proposition 2.6 que l'opérateur $((\nabla h + \lambda_k \partial f_X)^{-1} \circ \nabla h)$ est univoque, ce qui nous permet de conclure que la relation (3.6) est équivalente à

$$x^{k+1} = ((\nabla h + \lambda_k \partial f_X)^{-1} \circ \nabla h)(x^k) .$$

■

Ce qu'il fallait démontrer.

Il est maintenant temps de définir l'algorithme de minimisation proximale utilisant des fonctions de Bregman.

Algorithme 3.1 :

Soient $\{\lambda_k\}$ une suite de scalaires strictement positifs et h une fonction de Bregman de zone S telle que $D(\partial f_X) \subseteq S$.
Etant donné $x^0 \in S$, *l'algorithme de minimisation proximale utilisant des fonctions de Bregman* génère une suite $\{x^k\}_{k=0}^{\infty}$ telle que

$$x^{k+1} = \text{argmin}_{\,x \in X} \left\{ f(x) + \frac{1}{\lambda_k} D_h(x, x^k) \right\} . \tag{3.7}$$

Nous avons donc remplacé la résolution du problème (P_1) par celle d'une suite de problèmes d'optimisation ayant chacun une fonction objectif strictement convexe. En effet, il suit de la stricte convexité de la fonction h que la fonction D_h est elle aussi strictement convexe en son premier argument.

Ainsi, comme la conjuguée de Fenchel d'une fonction strictement convexe est différentiable, nous pouvons résoudre le sous-problème donnant x^{k+1} par dualité en utilisant les méthodes classiques de l'optimisation différentiable.

3.3 Convergence de l'algorithme de minimisation proximal utilisant des fonctions de Bregman

Nous énonçons les propriétés de convergence de l'algorithme de minimisation proximal utilisant des fonctions de Bregman dans le théorème suivant :

Théorème 3.2 :
Soient h une fonction de Bregman de zone S telle que $\overline{D(\partial f_X)} \subseteq S$, $\{\lambda_k\}$ une suite de scalaires strictement positifs telle que $\inf_{k \geq 0} \{\lambda_k\} > 0$ et $x^0 \in S$. Alors la suite $\{x^k\}_{k=0}$ définie par l'itération

$$x^{k+1} = \text{argmin}_{x \in X} \{f(x) + \frac{1}{\lambda_k} D_h(x, x^k)\}$$

converge vers un minimum de la fonction f sur l'ensemble X s'il en existe un. Dans le cas contraire, la suite $\{x^k\}$ est non bornée.

Preuve :

Le théorème 3.1 nous assure que

$$x^{k+1} = \text{argmin}_{x \in X} \left\{ f(x) + \frac{1}{\lambda_k} D_h(x, x^k) \right\}$$
$$\Leftrightarrow \quad x^{k+1} = \left((\nabla h + \lambda_k \partial f_X)^{-1} \circ \nabla h \right) (x^k) \, .$$

De plus, nous savons par hypothèse que $\overline{D(\partial f_X)} \subseteq S$.

Ainsi, il suit du théorème 2.7 que la suite $\{x^k\}$ converge vers un zéro de ∂f_X, c'est-à-dire un minimum de f sur X, s'il en existe un. Dans le cas contraire, la suite $\{x^k\}$ est non bornée. ∎

Il nous semble à présent utile d'établir la proposition suivante :

Proposition 3.3 :
Si $X \subseteq S$ alors $\overline{D(\partial f_X)} \subseteq S$.

Preuve :

Il suit de la proposition (A.IV.9) que $D(\partial f_X) \subseteq \text{dom } f_X$.
Or par hypothèse $\overline{\text{dom } f_X} = X \subseteq S$
Dès lors, nous obtenons $\overline{D(\partial f_X)} \subseteq X \subseteq S$. ■

Enfin, il nous semble intéressant de comparer notre analyse de convergence avec les travaux de Censor et Zenios (cfr [6]), et Gong Chen et Marc Teboulle (cfr [8]).

Remarque 3.1 :

L'itération (3.6) de l'algorithme de minimisation proximale avec fonctions de Bregman est identique à celle de l'algorithme P.M.D de Censor et Zenios (cfr [6]).

Cependant l'analyse présentée dans ce mémoire est légèrement plus générale.

En effet la fonction à minimiser f peut prendre la valeur $+\infty$ aussi longtemps qu'elle est s.c.i., alors que dans [6], la fonction f est à valeurs réelles.
De plus, notre étude de convergence ne nécessite pas que la fonction h possède des dérivées secondes continues, ni que la fonction D_h soit convexe en son second argument.
Finalement, le théorème 3.2 nous renseigne sur le comportement de l'algorithme en l'absence de solution.

Remarque 3.2 :

Comme nous l'avons signalé dans le paragraphe 2.4.4 (cfr remarque 2.2), l'hypothèse $\overline{D(T)} \subseteq S$, ou plus précisément $\overline{D(\partial f_X)} \subseteq S$, du théorème 3.2 est assez contraignante.

Dans leur article [8], Gong Chen et Marc Teboulle ont réussi à prouver la convergence de l'algorithme de minimisation proximale utilisant des fonctions de Bregman sous une hypothèse plus faible, en l'occurrence $ri(\text{dom } f_X) \subseteq S$.
De plus, ils ont pu exprimer l'ordre de convergence de l'algorithme en terme du résidu $f(x_k) - f(u)$ pour tout $u \in I\!R^n$.

Cependant leur analyse est différente de la nôtre dans le sens où elle est basée sur les propriétés des fonctions convexes et non plus sur les opérateurs maximaux monotones. Dès lors, ayant reçu leur article durant le mois de février, il ne nous a pas été possible d'étudier leur théorie.

3.4 Cas particulier : $h(x) = \frac{1}{2}||x||^2$

Nous désirons réécrire l'itération de l'algorithme de minimisation proximale utilisant la fonction de Bregman h de zone $I\!R^n$ définie par $h(x) = \frac{1}{2}||x||^2$.

Soit $x^0 \in I\!R^n$. Puisque $D_h(x,y) = \frac{1}{2}||x-y||^2 \ \forall\ x, y \in I\!R^n$, l'algorithme 3.1 génère une suite $\{x^k\}_{k=0}^{\infty}$ telle que

$$x^{k+1} = \text{argmin}_{\ x \in X}\ \{f(x) + \frac{1}{2\lambda_k}||x - x^k||^2\}$$

où $\{\lambda_k\}$ est une suite de scalaires strictement positifs.

De plus, comme $\nabla h = I\!R^n$ (cfr exemple 2.11), nous savons qu'une telle suite est bien définie, quel que soit le point de départ $x^0 \in I\!R^n$ (cfr assertion 3.2).

Nous retrouvons ainsi l'algorithme de minimisation proximale traditionnel développé par Martinet [15].

3.5 Résolution d'un problème de programmation linéaire

Dans ce paragraphe, nous désirons résoudre le problème de programmation linéaire suivant :

$$(PL) \qquad \begin{cases} \text{Minimiser} & < c, x > \\ \text{s.c.} & x \in X \end{cases}$$

où $X := \{x \in I\!R^n \text{ tel que } A_1x = b_1, A_2x \geq b_2, x \geq 0\}$, $c \in I\!R^n, b_1 \in I\!R^m, b_2 \in I\!R^p$, $A_1 \in I\!R^{m \times n}$ et $A_2 \in I\!R^{p \times n}$.

Comme l'a suggéré Eriksson [11], appliquons l'algorithme de minimisation proximale utilisant la fonction de Bregman $h(x) = -entx$ ou ent désigne la fonction entropique à ce problème.
Rappelons que, dans ce cas, la fonction h et la distance associée D_h sont définies respectivement sur $\overline{\Omega}^+$ et $\overline{\Omega}^+ \times \Omega^+$ par

$$h(x) = \sum_{i=1}^{n} x_i \ln \frac{x_i}{a_i}$$

où $a \in \overline{\Omega}^+$ et

$$D_h(x,y) = \sum_{i=1}^{n} x_i \left(\ln \frac{x_i}{y_i} - 1 \right) + \sum_{i=1}^{n} y_i \,.$$

Soient $\{\lambda_k\}$ une suite de scalaires strictement positifs et $x^0 \in \Omega^+$. L'itération de l'algorithme de minimisation proximale peut alors se réécrire

$$\begin{aligned} & x^{k+1} = \text{argmin}_{\ x \in X} \ \{f(x) + \frac{1}{\lambda_k} D_h(x, x^k)\} \\ \Leftrightarrow \quad & x^{k+1} = \text{argmin}_{\ x \in X} \left\{ \sum_{i=1}^{n} x_i c_i + \frac{1}{\lambda_k} \sum_{i=1}^{n} x_i \left(\ln \frac{x_i}{x_i^k} - 1 \right) + \frac{1}{\lambda_k} \sum_{i=1}^{n} x_i^k \right\} . \end{aligned}$$

Il s'agit d'un *problème d'optimisation purement entropique.* En effet, nous obtenons successivement que

$$\begin{aligned} & x^{k+1} = \text{argmin}_{\ x \in X} \left\{ \sum_{i=1}^{n} x_i \left(\ln e^{c_i} + \frac{1}{\lambda_k} \ln \frac{x_i}{x_i^k} - \frac{1}{\lambda_k} \ln e \right) \right\} \\ \Leftrightarrow \quad & x^{k+1} = \text{argmin}_{\ x \in X} \left\{ \frac{1}{\lambda_k} \sum_{i=1}^{n} x_i \ln \left(\frac{x_i}{x_i^k e} \quad (e^{c_i})^{\lambda_k} \right) \right\} . \end{aligned}$$

Posant alors $a_i = \frac{x_i^k e}{(e^{c_i})^{\lambda_k}}$ pour $i = 1, \cdots, n$, nous déduisons que

$$x^{k+1} = \text{argmax}_{\ x \in X} \left\{ \frac{-1}{\lambda_k} \sum_{i=1}^{n} x_i \ln \left(\frac{x_i}{a_i} \right) \right\} ,$$

et donc par la définition de fonction entropique

$$x^{k+1} = \text{argmax}_{\ x \in X} \ \frac{1}{\lambda_k} \ ent \ x \,.$$

Nous pouvons maintenant définir l'algorithme suivant :

Algorithme 3.2 :

Etant donné $x^0 \in \Omega^+$, *l'algorithme de minimisation proximale utilisant la fonction de Bregman* $h(x) = -ent\ x$ *appliqué au problème* (PL) génère une suite $\{x^k\}_{k=0}^{\infty}$ telle que

$$x^{k+1} = \text{argmax}_{\, x \in X} \left\{ \frac{1}{\lambda_k} ent\ x \right\} \tag{3.8}$$

où $\{\lambda_k\}$ est une suite de scalaires strictement positifs.

Notons que l'existence d'une telle suite est assurée par l'assertion (3.2). En effet, il suit de l'exemple 2.12 que $R(\nabla h) = I\!R^n$.

La *particularité* de cette méthode est qu'elle remplace la résolution d'un problème linéaire par celle d'une suite de problèmes non linéaires. De prime abord, ceci peut paraître assez étrange.

Cependant, nous disposons d'algorithmes "primal-dual" du type "action de ligne" très efficaces pour résoudre le problème purement entropique [cfr 5]. Nous pouvons donc les exploiter pour paralléliser l'itération (3.7). (Pour de plus amples renseignements, se référer à [7]).

Par conséquent, la résolution d'un problème linéaire par l'algorithme 3.2 implémenté sur des architectures de calcul en parallèle, pourra être plus *rapide* que celle obtenue par une méthode directe (comme par exemple l'algorithme du simplexe [9] où les méthodes de points intérieurs développées à partir de l'algorithme de Karmarkar [12]).

Chapitre 4

Méthode des multiplicateurs utilisant des fonctions de Bregman

Dans ce chapitre, nous désirons définir une nouvelle méthode non quadratique des multiplicateurs dans le cadre d'un problème de programmation convexe non différentiable.

Dans un premier temps, il nous semble utile de rappeler brièvement en quoi consiste cette méthode.

4.1 La méthode des multiplicateurs

La méthode des multiplicateurs combine la méthode de pénalité et la méthode "primal-dual". Elle est donc basée sur la résolution de problèmes d'optimisation sans contraintes. De plus, elle permet de résoudre le primal et le dual (au sens de Lagrange) d'un problème de programmation convexe.

Considérons tout d'abord les problèmes de minimisation convexe avec contraintes d'égalités, c'est-à-dire des problèmes de la forme :

$$(PE) \qquad \begin{cases} \text{Minimiser} & f(x) \\ \text{s.c.} & Ax = b \\ & x \in X \end{cases}$$

où f est une fonction propre, convexe et s.c.i., définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$; X est un ensemble convexe, fermé, non vide de $I\!R^n$; A est une matrice appartenant à $I\!R^{m\times n}$ de rang m et b est un vecteur de $I\!R^m$.
Nous supposons également que dom $f \cap \{x \in X$ tel que $Ax = b\} \neq \emptyset$ et que l'hypothèse de qualification de contraintes - $0 \in$ int $\{Ax - b$ tel que $x \in X\}$ - est vérifiée.

Définissant la fonction f_X de $I\!R^n$ dans $I\!R \cup \{+\infty\}$, par $f_X := f + \delta_X$, nous pouvons réécrire le problème (PE) sous la forme équivalente :

$$(\text{PE}_1) \qquad \begin{cases} \text{Minimiser} & f_X(x) \\ \text{s.c.} & Ax = b \end{cases}$$

La *fonction lagrangienne* associée à (PE) est donc définie de $I\!R^n \times I\!R^m$ dans $I\!R \cup \{+\infty\}$ par

$$L(x,v) = f_X(x) + < v, Ax - b > \ .$$

En outre, le vecteur v est appelé *multiplicateur de Lagrange* associé à (PE).

Dans la méthode classique des multiplicateurs proposée par Hestenes et Powell (1969), *le lagrangien augmenté* L_A est obtenu en ajoutant un terme de *pénalité quadratique* à la fonction lagrangienne du problème (PE). Plus précisément, la fonction L_A est définie de $I\!R^n \times I\!R^m \times I\!R_0^+$ dans $I\!R \cup \{+\infty\}$ par

$$L_A(x,v,c) = f_X(x) + < v, Ax - b > + \frac{1}{2} c ||Ax - b||^2 \ .$$

L'algorithme de *la méthode du lagrangien augmenté* pour (PE) est alors le suivant :

Algorithme 4.1 :

A l'itération k, soient un vecteur v^k de $I\!R^m$ et un scalaire c_k strictement positif connus.

1. Résoudre le problème de minimisation

$$x^k \in \text{argmin}_{\ x \in I\!R^n} \ \{L_A(x, v^k, c_k)\} \ .$$

2. Calculer l'itération des multiplicateurs

$$v^{k+1} = v^k + c_k(Ax^k - b) \ .$$

Intéressons-nous à présent aux problèmes de minimisation convexe avec contraintes d'inégalités, c'est-à-dire,

$$
\text{(PI)} \qquad \begin{cases} \text{Minimiser} & f(x) \\ \quad \text{s.c.} & g_i(x) \leq 0 \quad i = 1, \cdots, m \\ & x \in X \end{cases}
$$

où f et g_i $i = 1, \cdots, m$ sont des fonctions propres, convexes et s.c.i., définies de $I\!R^n$ dans $I\!R \cup \{+\infty\}$; et X est un sous-ensemble convexe, fermé, non vide de $I\!R^n$. Nous supposons également que les fonctions f et g_i $i = 1, \cdots, m$ ont une valeur finie sur X; que dom $f \cap \{x \in X \text{ tel que } g_i(x) \leq 0 \quad i = 1, \cdots, m\} \neq \emptyset$; et que l'hypothèse de qualification de contraintes - il existe $\hat{x} \in X$ tel que $g_i(\hat{x}) < 0 \quad i = 1, \cdots, m$ - est vérifiée.

Par la suite, nous noterons les contraintes d'inégalités $g(x) \leq 0$ où le vecteur $g(x)$ est défini par $g(x) = (g_1(x), \cdots g_m(x))^T$.

La fonction lagrangienne associée à (PI) est alors définie de $I\!R^n \times \bar{\Omega}^+$ dans $I\!R \cup \{+\infty\}$ par $L(x, u) = f_X(x) + < u, g(x) >$ où la fonction $f_X := f + \delta_X$ et $\bar{\Omega}^+ := \{u \in I\!R^m$ tel que $u \geq 0\}$.

Dans son article [18], Rockafellar a défini un *lagrangien augmenté quadratique* L_A analogue à celui de (PE). Plus précisément, L_A est défini de $I\!R^n \times \bar{\Omega}^+ \times I\!R_0^+$ dans $I\!R \cup \{+\infty\}$ par

$$L_A(x, u, c) = f_X(x) + \frac{1}{2c} \sum_{j=1}^{m} \{[\max\{0, u_j + c g_j(x)\}]^2 - u_j^2\} .$$

Enonçons maintenant l'algorithme de *la méthode du lagrangien augmenté* pour (PI).

Algorithme 4.2 :

A l'itération k soient u^k un vecteur de $\bar{\Omega}^+$ et c_k un scalaire strictement positif connus.

1. Résoudre le problème de minimisation

$$x^k \in \text{argmin}_{\, x \in I\!R^n} \{L_A(x, u^k, c_k)\} .$$

2. Calculer l'itération des multiplicateurs

$$u_i^{k+1} = \max\{0, u_i^k + c_k g_i(x^k)\} \quad \text{pour } i = 1 \cdots m .$$

Jusqu'à présent nous nous sommes intéressés à une fonction de pénalité quadratique. Cette fonction couramment utilisée, n'est toutefois pas appropriée à toutes les situations.

En effet, il se peut que le lagrangien augmenté quadratique L_A ne soit pas borné sur $\mathbb{R}^n$ alors que la fonction objectif est bornée inférieurement sur l'ensemble des contraintes [1, exemple (a), p. 302]. Dès lors, l'algorithme de minimisation du lagrangien augmenté diverge à moins que le point de départ ne soit proche de l'unique minimum local de L_A.

De plus, la fonction de pénalité influence fortement le taux de convergence de la méthode des multiplicateurs correspondante [1, p. 120]. Il est donc intéressant d'ajouter au lagrangien une fonction de pénalité adaptée au problème.

Dans son livre [1, chapitre 5], Bertsekas propose des fonctions de pénalité non quadratiques appropriées aux méthodes des multiplicateurs.
Citons à titre d'exemple *la fonction de pénalité exponentielle.* Cette fonction, utilisée pour des problèmes ne contenant que des contraintes d'inégalités, est définie par :

$$p(x,c) = c(e^x - 1) \ .$$

Dans ce cas, *le lagrangien augmenté* est une fonction définie de $\mathbb{R}^n \times \bar{\Omega}^+ \times \mathbb{R}_0^+$ dans $\mathbb{R} \cup \{+\infty\}$ par :

$$L_A(x,u,c) = f_X(x) + \frac{1}{c}\sum_{j=1}^{m} u^j(e^{cg_j(x)} - 1) \ ,$$

et la *méthode exponentielle des multiplicateurs* s'énonce :

Algorithme 4.3 :

A l'itération k, soient u^k un vecteur de $\bar{\Omega}^+$ et c_k un scalaire strictement positif connus.

1. Résoudre le problème de minimisation

$$x^k \in \text{argmin}_{\, x\in\mathbb{R}^n} \ L_A(x, u^k, c_k) \ .$$

2. Calculer l'itération des multiplicateurs

$$u_i^{k+1} = u_i^k e^{c_k g_i(x^k)} \quad i = 1 \cdots m \ .$$

Pour terminer, remarquons que la méthode des multiplicateurs élimine partiellement les désavantages de l'approche de pénalité et de l'approche "primal-dual".

En effet, la convergence de la méthode des multiplicateurs ne nécessite pas de faire tendre le paramètre de pénalisation vers l'infini. On évite ainsi le mauvais conditionnement qui altère la méthode de pénalité (où le terme de pénalité est directement ajouté à la fonction objectif).

De plus, dans le cas des fonctions de pénalité quadratiques, l'itération des multiplicateurs tend à converger vers une solution du problème dual (au sens de Lagrange) plus vite que dans la méthode "primal-dual" (où x^k est obtenu en minimisant la fonction lagrangienne traditionnelle associée au problème).

4.2 Motivation

Comme l'a observé Rockafellar [17, p. 107], le lagrangien augmenté quadratique des problèmes de minimisation convexe avec contraintes d'égalités (PE), ou avec contraintes d'inégalités (PI), est *l'approximation de Moreau-Yosida* de la fonction lagrangienne. Plus précisément, pour tout $c > 0, x \in I\!R^n$ et $p \in I\!R^m$, nous avons que

$$L_A(x,p,c) = \max_{\xi \in I\!R^m} \{L(x,\xi) - \frac{1}{2c}||\xi - p||^2\} .$$

Rappelons que dans le cas de (PI), le multiplicateur de lagrange est positif. Dans ce cas, la maximisation se fait sur les $\xi \in I\!R^m$ tel que $\xi \geq 0$.

Nous pouvons donc définir un lagrangien augmenté généralisé L_B en remplaçant le terme quadratique de l'approximation de Moreau-Yosida par un terme incluant la fonction distance D_h associée à une fonctionde Bregman h. Plus précisément L_B est défini de $I\!R^n \times I\!R^m \times I\!R_0^+$ par

$$L_B(x,p,c) = \sup_{\xi \in I\!R^m} \{L(x,\xi) - \frac{1}{c}D_h(\xi,p)\} .$$

De nouveau, s'il s'agit d'un problème de la forme (PI), nous imposons que ξ soit positif.

Par ailleurs, comme nous le constaterons dans les paragraphes (4.3.1 et 4.4.1), le problème dual (au sens de Lagrange) d'un problème de minimisation convexe, soumis à des contraintes d'égalités ou d'inégalités, est équivalent à la recherche d'un zéro de l' opérateur maximal monotone $\partial(-d)$, (où $\partial(-d)$ désigne le sous-différentiel de l'opposé de la fonction duale (au sens de Lagrange) du problème considéré).

Ainsi, il suffit d'appliquer l'algorithme du point proximal utilisant des fonctions de Bregman (algorithme 2.1) à l'opérateur $\partial(-d)$, pour obtenir la suite de multiplicateurs $\{p^k\}$.

Il est maintenant temps d'énoncer l'algorithme de *la méthode des multiplicateurs utilisant des fonctions de Bregman.*

Algorithme 4.4 :

Soit h une fonction de Bregman de zone $S \subseteq I\!R^m$ telle que $D(\partial(-d)) \subseteq S$.
A l'itération k, soient p^k un vecteur de $I\!R^m$ et λ_k un scalaire strictement positif connus

1. Résoudre le problème de minimisation

$$x^k \in \operatorname{argmin}_{x \in I\!R^n} \ \{L_B(x, p, \lambda_k)\}$$

où $L_B(x, p, \lambda_k) := \sup_{\xi \in I\!R^m} \{L(x, \xi) - \frac{1}{\lambda_k} D_h(\xi, p)\}$.

2. Calculer l'itération des multiplicateurs

$$p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k) \ .$$

4.3 Problème de minimisation convexe avec contraintes d'égalités

Dans ce paragraphe, nous désirons particulariser la méthode des multiplicateurs utilisant des fonctions de Bregman au problème de minimisation convexe ne contenant que des contraintes d'égalités. Pour rappel

$$(PE) \qquad \begin{cases} \text{Minimiser} & f(x) \\ \text{s.c.} & Ax = b \\ & x \in X \end{cases}$$

où f est une fonction propre, convexe et s.c.i., définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$; X est un ensemble convexe, fermé, non vide de $I\!R^n$; A est une matrice appartenant à $I\!R^{m \times n}$ de rang m et b est un vecteur de $I\!R^m$.
De plus nous supposons que :

- l'ensemble dom $f \cap \{x \in X$ tel que $Ax = b\}$ est non vide
- l'hypothèse de qualification de contraintes - $0 \in \text{int}\ \{Ax - b$ tel que $x \in X\}$ - est vérifiée.

Dans un premier temps, nous exprimerons le problème dual (au sens de Lagrange) du problème (PE) en terme de la recherche d'un zéro d'un opérateur maximal monotone (cfr paragraphe 4.3.1).

Dans un second temps, nous adapterons l'algorithme de la méthode des multiplicateurs utilisant des fonctions de Bregman au problème (PE) (cfr paragraphe 4.3.2).

Ensuite, nous proposerons des conditions d'existence et d'unicité des itérés, et nous analyserons la convergence de l'algorithme adapté à (PE) (cfr paragraphes 4.3.3 et 4.3.4).

Pour terminer, nous vérifierons que par le choix de la fonction de Bregman $h(x) = \frac{1}{2}||x||^2$, nous retrouvons la méthode quadratique du lagrangien augmenté (cfr paragraphe 4.3.5).

4.3.1 Le dual au sens de Lagrange du problème (PE)

La *fonction duale Lagrangienne* du problème (PE) est définie de $I\!R^m$ dans $I\!R \cup \{-\infty\}$ par

$$d(p) = \inf_{x \in I\!R^n} \{f_X(x) + < p, Ax - b >\}, \tag{4.1}$$

c'est-à-dire

$$d(p) = \inf_{x \in I\!R^n} \{L(x,p)\}$$

où $L(x,p)$ est la fonction Lagrangienne associée au problème (PE).

Il suit de l'hypothèse de qualification de contraintes que la fonction duale est propre. De plus, puisque la fonction $-d$ est convexe et s.c.i., nous obtenons que la fonction d est concave et s.c.i. [19, p. 307-p. 308].

Enonçons maintenant le *dual au sens de Lagrange* du problème (PE)

$$\text{(DE)} \qquad \begin{cases} \text{Maximiser} & d(p) \\ \text{s.c.} & p \in I\!R^m \end{cases}$$

Une solution p^* du problème (DE) est appelée *multiplicateur de Lagrange optimal* pour (PE).

En vertu de la condition d'optimalité de (DE)

$$0 \in \partial(-d(p^*)),$$

nous concluons que *la résolution du problème (DE) est équivalente à la recherche d'un zéro de l'opérateur $\partial(-d)$*. Ce dernier est *maximal monotone* puisque la fonction $-d$ est propre, convexe et s.c.i. (cfr exemple 2.5).

4.3.2 Algorithme de la méthode des multiplicateurs utilisant des fonctions de Bregman pour le problème (PE)

Soient h une fonction de Bregman de zone $S \subseteq I\!R^m$ telle que $D(\partial(-d)) \subseteq S$, $\{\lambda_k\}$ une suite de scalaires strictement positifs et p^0 un vecteur de $I\!R^m$. Appliquant l'algorithme de la méhode des multiplicateurs au problème (PE), nous obtenons une suite $\{(x^k, p^k)\}_{k=0}^{\infty} \subseteq I\!R^n \times I\!R^m$ telle que

$$\begin{aligned} x^k &\in \operatorname{argmin}_{x \in I\!R^n} \{L_B(x, p^k, \lambda_k)\} && (4.2) \\ p^{k+1} &= ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k) && (4.3) \end{aligned}$$

où le lagrangien augmenté généralisé L_B est défini par

$$L_B(x, p^k, \lambda_k) = \sup_{\xi \in I\!R^m} \{L(x, \xi) - \frac{1}{\lambda_k} D_h(\xi, p^k)\} .$$

Avant d'adapter la relation d'itération de la suite $\{(x^k, p^k)\}$ au problème (PE), rappelons la définition et quelques propriétés de la conjuguée de Fenchel d'une fonction convexe.

Définition 4.1 : Conjuguée de Fenchel.
Soit f une fonction convexe définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$. Alors la fonction f^* définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$ par

$$f^*(y) = \sup_{x \in I\!R^n} \{< x, y > -f(x)\}$$

est appelée la *fonction conjuguée de Fenchel* de f.

Proposition 4.1 :
Si la fonction f est convexe et propre, alors la fonction conjuguée de f, ie f^*, est propre, convexe et s.c.i.

Preuve : cfr [19], théorème 12.2, p. 104.

Proposition 4.2 :
Si f est une fonction propre, convexe et fermée, alors ∂f^* est l'inverse de ∂f dans le sens des opérateurs multivoques, c'est-à-dire

$$x \in \partial f^*(x^*) \Leftrightarrow x^* \in \partial f(x) .$$

Preuve : cfr [19], corollaire 23.5.1, p. 219.

Réécrivons à présent les relations d'itération (4.2) et (4.3) en terme des fonctions f_X, h et h^*.

Théorème 4.1 : *Forme de l'itération*
Soient h une fonction de Bregman de zone $I\!R^m$ telle que $R(\nabla h) = I\!R^m$, $\{\lambda_k\}$ une suite de scalaires strictement positifs, et $p^0 \in I\!R^m$.
Considérons la suite $\{(x^k, p^k)\} \subseteq X \times I\!R^m$ définie par

$$\begin{cases} x^k \in \operatorname{argmin}_{x \in I\!R^n} \; \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))\} \\ p^{k+1} = \nabla h^*(\nabla h(p^k) + \lambda_k(Ax^k - b)) \; . \end{cases}$$

Alors nous obtenons que

$$\begin{cases} x^k \in \operatorname{argmin}_{x \in I\!R^n} \; \{L_B(x, p^k, \lambda_k)\} \\ p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k) \; . \end{cases}$$

Dans le but de prouver ce résultat, rappelons quelques notions théoriques relatives aux fonctions essentiellement régulières et montrons quelques lemmes techniques.

Définition 4.2 :
Soit f une fonction propre, convexe, définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$.
Si f vérifie les trois conditions suivantes :

(a) int (dom f) $\neq \emptyset$

(b) f est différentiable sur int (dom f)

(c) $\lim_{i \to \infty} |\nabla f(x_i)| = +\infty$ où la suite de vecteurs $\{x_i\} \subseteq$ int (dom f) converge vers un vecteur x de la frontière de int (dom f).

alors f est *essentiellement régulière.*

Proposition 4.3 :
Soit f une fonction propre, convexe et fermée. Alors l'opérateur ∂f est univoque si et

seulement si f est essentiellement régulière. De plus, dans ce cas

$$\partial f(x) = \begin{cases} \nabla f(x) & \text{si } x \in \text{ int (dom } f) \\ \emptyset & \text{sinon.} \end{cases}$$

__Preuve :__ cfr [19], théorème 26.1, p. 251.

__Définition 4.3 :__
Une fonction f propre sur $I\!R^n$ est *essentiellement strictement convexe* si f est strictement convexe sur tout sous-ensemble convexe de $D(\partial f)$.

__Proposition 4.4 :__
Une fonction propre, convexe, fermée est essentiellement strictement convexe si et seulement si sa conjuguée est essentiellement régulière.

__Preuve :__ cfr [19], théorème 26.3, p. 253.

__Lemme 4.1 :__
Soit h une fonction de Bregman de zone $S \subseteq I\!R^m$. Alors la fonction h^* est différentiable sur int (dom h^*).

__Preuve du lemme 4.1 :__

La fonction h est essentiellement strictement convexe sur S. En effet, il suit de la définition d'une fonction de Bregman (définition 2.11) que h est strictement convexe et différentiable sur S. La thèse découle alors de la définition 4.3.

De plus, la fonction h étant propre, convexe et continue, elle est fermée (cfr proposition A.III.4).

Nous déduisons donc de la proposition 4.4 que la conjuguée de h, ie h^*, est une fonction essentiellement régulière.

Dès lors, la proposition 4.3 nous affirme que la fonction h^* est différentiable sur int (dom h^*).
Ce qu'il fallait démontrer. ■

Lemme 4.2 :
Soient h une fonction de Bregman de zone $I\!R^m$ telle que $R(\nabla h) = I\!R^m$, et $[\lambda_k\}$ une suite de scalaires strictement positifs. Etant donné $p^0 \in I\!R^m$, considérons la suite $\{p^k\}_{k=0}^{\infty} \subseteq I\!R^m$ définie par

$$p^{k+1} = \nabla h^*(\nabla h(p^k) + \lambda_k(Ax^k - b)) .$$

Alors les trois assertions suivantes sont équivalentes :

1. $x^k \in \text{argmin}_{\ x \in I\!R^n} \ \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))\}$
2. $-A^T p^{k+1} \in \partial f_X(x^k)$
3. $x^k \in \partial (f_X)^*(-A^T p^{k+1})$

Preuve du lemme 4.2 :

a) Montrons l'équivalence des assertions (1.) et (2.).

- Remarquons dès à présent que dom $h^* = I\!R^m$. En effet, il suit de la proposition A.IV.9 que $R(\nabla h) \subseteq$dom $h^* \subseteq I\!R^m$. Or par hypothèse $R(\nabla h) = I\!R^m$, ainsi nous obtenons que dom $h^* = I\!R^m$.
- Comme

$$x^k \in \text{argmin}_{\ x \in I\!R^n} \ \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))\} ,$$

il suit que

$$0 \in \partial[f_X(.) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(A. - b))](x^k) . \tag{4.6}$$

Considérons alors la fonction définie sur $I\!R^n$ par

$$f_1(x) = \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b)) .$$

Il est évident que $ri(\text{dom } f_1) \cap ri(\text{dom } f_X) \neq \emptyset$. En effet, d'une part dom $f_1 = I\!R^m$ car dom $h^* = I\!R^m$. D'autre part $ri(f_X) = ri(x) \neq \emptyset$ car $X \subseteq I\!R^n$ est un ensemble convexe (cfr proposition A.II.4).

Ainsi, nous pouvons réécrire la relation (4.6) de manière équivalente

$$0 \in \partial f_X(x^k) + \partial[h^*(\nabla h(p^k) + \lambda_k(A. - b))](x^k) ,$$

ou encore, par la différentiabilité de la fonction h^* sur int (dom h^*) $= I\!R^m$ (cfr lemme 4.1)

$$0 \in \partial f_X(x^k) + \nabla[h^*(\nabla h(p^k) + \lambda_k(A. - b))](x^k) . \tag{4.7}$$

Or $R(A) \cap ri(\text{dom } h^*) \neq \emptyset$ car A est une transformation linéaire de rang m et dom $h^* = I\!R^m$. Par conséquent, la proposition A.IV.13 nous assure que

$$\nabla[h^*(\nabla h(p^k) + \lambda_k(A. - b))](x^k) = A^T \nabla h^*(\nabla h(p^k) + \lambda_k(A. - b))(x^k) .$$

Dès lors, la relation (4.7) est équivalente à

$$0 \in \partial f_X(x^k) + A^T \nabla h^*(\nabla h(p^k) + \lambda_k(Ax^k - b)) ,$$

c'est-à-dire par la définition de p^{k+1},

$$0 \in \partial f_X(x^k) + A^T p^{k+1} ,$$

ou encore

$$-A^T p^{k+1} \in \partial f_X(x^k) .$$

b) L'équivalence des assertions (2.) et (3.) suit immédiatement du fait que les opérateurs ∂f_X et $\partial(f_X)^*$ sont deux opérateurs inverses (cfr proposition 4.2). ■

Lemme 4.3 :

Soit d la fonction duale (au sens de Lagrange) du problème (PE). Alors

$$\forall\, p \in I\!R^m \quad \partial(-d)(p) = \partial[(f_X)^* \circ (-A^T.)](p) + b$$

Preuve du lemme 4.3 :

Il suit des définitions de la fonction duale et de la conjuguée de Fenchel que pour tout $p \in I\!R^m$,

$$\begin{aligned} & -d(p) = -\inf{}_{x \in I\!R^*} \ \{f_X(x) + < p, Ax - b >\} \\ \Leftrightarrow\ & -d(p) = \sup_{x \in I\!R^n} \ \{< -A^T p, x > - f_X(x)\} + < b, p > \\ \Leftrightarrow\ & -d(p) = (f_X)^*(-A^T p) + < b, p > \end{aligned}$$

Ainsi nous obtenons que pour tout $p \in I\!R^m$

$$\partial(-d)(p) = \partial_p[(f_X)^* \circ (-A^T.) + < b, . >](p) \tag{4.8}$$

où ∂_p désigne l'ensemble des sous-gradients par rapport à la coordonnée p.

Définissons à présent les fonctions f_1 et f_2 sur $I\!R^m$ par $f_1(p) =< b, p >$ et $f_2(p) = ((f_X)^* \circ (-A^T))(p)$.

Remarquons que $ri(\text{dom } f_1) \cap ri(\text{dom } f_2) \neq \emptyset$. En effet, il est évident que dom $f_1 = I\!R^m$. De plus, l'ensemble $ri(\text{dom } f_2) \subseteq I\!R^m$ est non vide car dom f_2 est convexe (cfr proposition A.II.4).

Par conséquent le théorème relatif au sous-différentiel d'une somme (cfr proposition A.IV.12) nous assure que la relation (4.8) est équivalente à

$$\partial(-d)(p) = \partial_p[(f_X)^*(-A^T.)](p) + \nabla_p(< b, . >)(p)$$

La thèse suit alors de la définition du gradient du produit scalaire. ∎

Nous disposons maintenant de tous les résultats nécessaires pour prouver le théorème relatif à la forme de l'itération de la méthode des multiplicateurs utilisant des fonctions de Bregman pour le problème (PE).

<u>**Preuve du théorème 4.1 :**</u>

<u>Pas 1 :</u> Prouvons d'abord que

$$x^k \in \text{argmin}_{\,x \in I\!R^n} \; \{L_B(x, p^k, \lambda_k)\} \Leftrightarrow x^k \in \text{argmin}_{\,x \in I\!R^n} \; \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))\} \, .$$

Comme

$$L_B(x, p^k, \lambda_k) = \sup_{\xi \in I\!R^m} \{L(x, \xi) - \frac{1}{\lambda_k} D_h(\xi, p^k)\} \, ,$$

il suit des définitions du lagrangien associé à (PE) et de la fonction D_h que

$$\begin{aligned} L_B(x, p^k, \lambda_k) &= \sup_{\xi \in I\!R^m} \{f_X(x) + < \xi, Ax - b > - \frac{1}{\lambda_k} h(\xi) + \frac{1}{\lambda_k} h(p^k) \\ &+ \frac{1}{\lambda_k} < \nabla h(p^k), \xi - p^k >\} \, , \end{aligned}$$

c'est-à-dire,

$$\begin{aligned} L_B(x, p^k, \lambda_k) &= f_X(x) + \frac{1}{\lambda_k} h(p^k) - \frac{1}{\lambda_k} < \nabla h(p^k), p^k > \\ &+ \lambda_k \sup_{\xi \in I\!R^m} \{< \nabla h(p^k) + \lambda_k(Ax - b), \xi > - h(\xi)\} \, . \end{aligned}$$

Par conséquent, la définition de la conjuguée de Fenchel de h (définition 4.1) nous assure que

$$\begin{aligned} L_B(x, p^k, \lambda_k) &= f_X(x) + \frac{1}{\lambda_k} h(p^k) - \frac{1}{\lambda_k} < \nabla h(p^k), p^k > \\ &+ \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b)) \, . \end{aligned}$$

La thèse suit alors de l'indépendance en x du terme $\frac{1}{\lambda_k}h(p^k) - \frac{1}{\lambda_k} < \nabla h(p^k), p^k >$.

Pas 2 : Montrons à présent que
si $p^{k+1} = \nabla h^*(\nabla h(p^k) + \lambda_k(Ax^k - b))$ alors $p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k)$.

- Il suit d'abord de la définition de p^{k+1} que

$$\frac{1}{\lambda_k}[\nabla h(p^k) - \nabla h(p^{k+1})] = \frac{1}{\lambda_k}\{\nabla h(p^k) - \nabla h(\nabla h^*(\nabla h(p^k) + \lambda_k(Ax^k - b)))\} .$$

De plus ∇h et ∇h^* sont des opérateurs inverses car h est une fonction propre, convexe et continue sur $\mathbb{R}^m$ (cfr proposition 4.2). Ainsi, nous obtenons que

$$\frac{1}{\lambda_k}[\nabla h(p^k) - \nabla h(p^{k+1})] = -(Ax^k - b) . \tag{4.9}$$

- D'autre part, le lemme 4.2 ($1 \Leftrightarrow 3$) nous assure que

$$x^k \in \partial(f_X)^*(-A^T p^{k+1}) .$$

Or il suit du théorème (A.IV.13) que

$$\partial((f_X)^* \circ (-A^T)) \subseteq (-A) \circ \partial((f_X)^* \circ (-A^T)) ,$$

dès lors

$$- Ax^k \in \partial_p[(f_X) \circ (-A^T p)](p^{k+1}) \tag{4.10}$$

où ∂_p désigne l'ensemble des sous-gradients par rapport à la coordonnée p.

- Par conséquent, nous déduisons des relations (4.9) et (4.10) que

$$\frac{1}{\lambda_k}(\nabla h(p^k) - \nabla h(p^{k+1})) \in \partial_p[(f_X)^* \circ (-A^T p)](p^{k+1}) + b .$$

Ainsi, comme par le lemme 4.3

$$\partial(-d)(p^{k+1}) = \partial_p[(f_X)^* \circ (-A^T p)](p^{k+1}) + b ,$$

nous obtenons que

$$\frac{1}{\lambda_k}(\nabla h(p^k) - \nabla h(p^{k+1})) \in \partial(-d)(p^{k+1}) ,$$

c'est-à-dire

$$p^{k+1} \in ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k)$$

ou de façon équivalente,

$$p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k) .$$

En effet, comme l'opérateur $\partial(-d)$ est monotone il suit alors de la proposition 2.7 que l'opérateur $((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)$ est univoque. ∎

Il est maintenant temps d'énoncer l'algorithme de *la méthode des multiplicateurs utilisant des fonctions de Bregman dans le cas d'un problème de la forme* (PE).

Algorithme 4.5 :

Soient $\{\lambda_k\}_{k=0}^{\infty}$ une suite de scalaires strictement positifs et h une fonction de Bregman de zone $\mathbb{R}^m$ telle que $R(\nabla h) = \mathbb{R}^m$. A l'itération k, soient p^k un vecteur de $\mathbb{R}^m$ et λ_k un scalaire strictement positif connus.

1. Résoudre le problème de minimisation

$$x^k \in \operatorname{argmin}_{x \in \mathbb{R}^n} \ \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))\} .$$

2. Calculer l'itération des multiplicateurs

$$p^{k+1} = \nabla h^*(\nabla h(p^k) + \lambda_k(Ax^k - b)) .$$

Notons que par le changement de variable $\mu^k = \nabla h(p^k)$ pour tout $k \geq 0$, les itérés de l'algorithme 4.5 peuvent se réécrire

$$\left\{ \begin{array}{l} x^k \in \operatorname{argmin}_{x \in \mathbb{R}^n} \ \{f_X(x) + \frac{1}{\lambda_k} h^*(\mu^k + \lambda_k(Ax - b))\} \\ \mu^{k+1} = \mu^k + \lambda_k(Ax^k - b)\} \end{array} \right. .$$

La formule de mise à jour des multiplicateurs est donc de forme semblable à celle de la méthode du lagrangien augmenté quadratique (cfr algorithme 4.1).

4.3.3 Conditions d'existence et d'unicité des itérés

Intéressons nous d'abord aux conditions **d'existence de la suite** $\{(x^k, p^k)\}_{k=0}^{\infty}$ générée par la méthode des multiplicateurs avec fonctions de Bregman adaptée à (PE).

Théorème 4.2 :

Soient h une fonction de Bregman de zone $I\!R^m$ telle que $R(\nabla h) = I\!R^m$ et $\{\lambda_k\}$ une suite de scalaires strictement positifs.

Supposons également que l'une des deux hypothèses suivantes est vérifiée :

(H.4.1) l'espace engendré par les lignes de A intersecte $ri(R(\partial f_X))$

(H.4.2) la fonction f et l'ensemble X sont polyhédrals et l'espace engendré par les lignes de A intersecte $R(\partial f_X)$.

Alors, la suite $\{(x^k, p^k)\}$ générée par l'itération

$$\begin{cases} x^k \in \text{argmin}_{x \in I\!R^n} \ \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))\} \\ p^{k+1} = \nabla h^*(\nabla h(p^k) + \lambda_k(Ax^k - b)) \end{cases}$$

est bien définie quel que soit le point de départ $p^0 \in I\!R^m$.

Avant de prouver ce résultat, rappelons la définition d'un ensemble et d'une fonction polyhédrals.

Définition 4.4 :

Un *ensemble* de $I\!R^n$ est *polyhédral* convexe s'il peut s'exprimer comme l'intersection d'une collection finie de demi-espaces fermée de $I\!R^n$.

Définition 4.5 :

Soit f une fonction convexe définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$. Si l'épigraphe de f est un ensemble polyhédral convexe de $I\!R^{n+1}$ alors f est une *fonction polyhédrale* convexe sur $I\!R^n$.

Exemple 4.1 :

Si X est un ensemble polyhédral convexe de $I\!R^n$ alors la fonction indicatrice de X, c'est-à-dire δ_X, est une fonction polyhédrale convexe.

Proposition 4.5 :
Si f_1 et f_2 sont des fonctions polyhédrales convexes sur $I\!R^n$, alors $f_1 + f_2$ est une fonction polyhédrale convexe.

Preuve : Cfr [19], théorème 19.4, p. 176.

Prouvons maintenant le théorème relatif à l'existence des itérés de la méthode des multiplicateurs avec fonctions de Bregman.

Preuve du théorème 4.2 :

Pas 1 : Montrons d'abord l'existence de la suite $\{p^k\}$.

Nous savons par le théorème 4.1 que $p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k)$. Or par hypothèse $R(\nabla h) = I\!R^m$. Ainsi, le théorème 2.6, appliqué à l'opérateur maximal monotone $\partial(-d)$, nous assure que la suite $\{p^k\}_{k=0}^{\infty}$ existe pour tout $p^0 \in I\!R^m$.

Pas 2 : Intéressons-nous à présent à l'existence de la suite $\{x^k\}$.

Dans un premier temps nous montrerons l'équivalence suivante :

$$x^k \in \operatorname{argmin}_{x \in I\!R^n} \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + (Ax - b))\} \Leftrightarrow -Ax^k \in \partial((f_X)^* \circ (-A^T p))(p^{k+1})\ .$$

Ensuite, nous vérifierons que pour tout $k \geq 0$, il existe $x^k \in I\!R^n$ tel que

$$-Ax^k \in \partial((f_X)^* \circ (-A^T p))(p^{k+1})\ .$$

Ce qui nous permettra de conclure que pour tout $k \geq 0$, il existe x^k tel que

$$x^k \in \operatorname{argmin}_{x \in I\!R^n} \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))\}\ .$$

a) Montrons d'abord que

$$\begin{aligned} & x^k \in \operatorname{argmin}_{x \in I\!R^n} \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))\} \\ \Leftrightarrow \quad & -Ax^k \in \partial((f_X)^* \circ (-A^T p))(p^{k+1})\ . \end{aligned} \tag{4.11}$$

- D'une part il résulte du lemme 4.2 ($1 \Leftrightarrow 3$) que

$$\begin{aligned} & x^k \in \operatorname{argmin}_{x \in I\!R^n} \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))\} \\ \Leftrightarrow \quad & x^k \in \partial(f_X)^* \circ (-A^T p^{k+1})\ . \end{aligned} \tag{4.12}$$

- D'autre part, les hypothèses (H.4.1) et (H.4.2) sont nécessaires et suffisantes pour que
$$\partial((f_X)^* \circ (-A^T)) = (-A) \circ \partial(f_X^* \circ (-A^T)) \,. \tag{4.13}$$
En effet

(i) Supposons d'abord que la condition (H.4.1) est vérifiée, c'est-à-dire :
$$\{-A^T x \text{ tel que } x \in I\!R^n\} \cap ri(R(\partial f_X)) \neq \emptyset \,.$$

- Dans ce cas
$$ri(R(\partial f_X)) \subseteq ri(\text{dom } (f_X)^*) \tag{4.14}$$
En effet, la fonction f_X étant propre et convexe, il suit des propositions A.IV.9 et A.II.9 que
$$ri(\text{dom } (f_X)^*) \subseteq R(\partial f_X) \subseteq \text{dom } (f_X)^* \tag{4.15}$$
et
$$\dim\ (ri(\text{dom } (f_X)^*)) = \dim\ (\text{dom } (f_X)^*)$$
où dim (A) désigne la dimension de l'ensemble A.
Ainsi, nous obtenons que
$$\dim\ (\text{dom } (f_X)^*) = \dim\ (R(\partial f_X))\,,$$
d'où, par la définition de la dimension d'un ensemble convexe [cfr 19, p. 12],
$$\dim\ (aff(\text{dom } (f_X)^*)) = \dim\ (aff(R(\partial f_X)))\,. \tag{4.16}$$
Par ailleurs, il suit de l'inclusion $R(\partial f_X) \subset \text{dom } (f_X)^*$ (cfr (4.15)) que
$$aff(R(\partial f_X)) \subseteq aff(\text{dom } (f_X)^*)\,. \tag{4.17}$$
Dès lors, nous déduisons des relations (4.16) et (4.17) que
$$aff(R(\partial f_X)) = aff(\text{dom } (f_X)^*)\,,$$
or par la relation $R(\partial f_X) \subseteq \text{dom } (f_X)^*$,
ainsi, il résulte de la définition de l'intérieur relatif (cfr A.II.2) que
$$ri(R(\partial f_X)) \subseteq ri(\text{dom } (f_X)^*)$$

- Par conséquent, puisque par l'hypothèse H.4.1 :

$$\{-A^T x \text{ tel que } x \in I\!R^n\} \cap ri(R(\partial f_X)) \neq \emptyset$$

et par la relation (4.14)

$$ri(R(\partial f_X)) \subseteq ri(\text{dom } (f_X)^*) ,$$

nous concluons que

$$\{-A^T x \text{ tel que } x \in I\!R^n\} \cap ri(\text{dom } (f_X)^*) \neq \emptyset .$$

La thèse résulte alors du théorème A.IV.13.

(ii) Supposons maintenant que la condition (H.4.2) est vérifiée, c'est-à-dire f et X sont polyhédrals et $\{-A^T x \text{ tel que } x \in I\!R^n\} \cap R(\partial f_X) \neq \emptyset$.

Dans ce cas, il suit de l'exemple 4.1 et de la proposition 4.5 que la fonction $f_X := f + \delta_X$ est une fonction polyhédrale.
Par ailleurs il résulte de l'assertion A.IV.9

$$R(\partial f_X) \subseteq \text{dom } f_X .$$

Dès lors, l'hypothèse H.4.2 nous assure que la fonction f_X est polyhédrale et que

$$\{-A^T x \text{ tel que } x \in I\!R^n\} \cap \text{ dom } (f_X)^* \neq \emptyset .$$

De nouveau, la thèse découle du théorème (A.IV.13).

- Finalement, nous déduisons de (4.12) et (4.13) que

$$\begin{aligned} & x^k \in \operatorname{argmin}_{x \in I\!R^n} \ \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))\} \\ \Leftrightarrow \quad & x^k \in \partial(f_X)^*(-A^T p^{k+1}) \\ \Leftrightarrow \quad & -Ax^k \in \partial((f_X)^* \circ (-A^T p))(p^{k+1}) . \end{aligned}$$

b) Montrons maintenant que pour tout $k \geq 0$, il existe $x^k \in I\!R^n$ tel que

$$-Ax^k \in \partial((f_X)^* \circ (-A^T p))(p^{k+1}) . \tag{4.18}$$

- Nous savons par le lemme 4.3 que pour tout $k \geq 0$,

$$\partial((f_X)^* \circ (-A^T p))(p^{k+1}) + b = \partial(-d)(p^{k+1}) . \tag{4.19}$$

- D'autre part $\partial(-d)(p^{k+1}) \neq \emptyset$ pour tout $k \geq 0$. En effet, le pas 1 nous assure que pour tout $k \geq 0$, il existe p^{k+1} tel que

$$p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k) .$$

Autrement dit, pour tout $k \geq 0$, il existe p^{k+1} tel que

$$\frac{1}{\lambda_k}(\nabla h(p^k) - \nabla h(p^{k+1})) \in \partial(-d)(p^{k+1}) . \tag{4.20}$$

- Dès lors il suit des relations (4.19) et (4.20) que pour tout $k \geq 0$

$$\partial((f_X)^* \circ (-A^T p))(p^{k+1}) \neq \emptyset .$$

Ainsi, puisque par hypothèse la matrice A est de rang m, nous obtenons que pour tout $k \geq 0$, il existe $x^k \in \mathbb{R}^n$ tel que

$$-Ax^k \in \partial((f_X)^* \circ (-A^T p))(p^{k+1}) .$$

c) En conclusion, nous déduisons des relations (4.11) et (4.18) que pour tout $k \geq 0$, il existe x^k tel que

$$x^k \in \text{argmin}_{x \in \mathbb{R}^n} \ \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))\} .$$

Ce qu'il fallait démontrer. ■

Pour terminer, remarquons que le caractère univoque de l'opérateur $((\nabla h + \lambda_k \partial(-d))^{-1} \nabla h)$ assure **l'unicité de** p^k pour tout $k \geq 0$.

Par contre, nous ne pouvons, en général, rien affirmer au sujet de **l'unicité de** x^k. Toutefois, lorsque la fonction f est strictement convexe alors la fonction définie sur $\mathbb{R}^n$ par $f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))$ est elle aussi strictement convexe.
Dès lors il existe un seul $x^k \in \mathbb{R}^n$ tel que

$$x^k = \text{argmin}_{x \in \mathbb{R}^n} \ \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax - b))\} .$$

4.3.4 Convergence vers une solution des problèmes primal (PE) et dual (DE)

Enonçons directement le théorème de convergence de la méthode des multiplicateurs utilisant des fonctions de Bregman vers une solution des problèmes (PE) et (DE).

<u>Théorème 4.3 :</u>
Soient h une fonction de Bregman de zone $I\!R^m$ telle que $R(\nabla h) = I\!R^m$, et $\{\lambda_k\}_{k=0}^{\infty}$ une suite de scalaires strictement positifs tel que $\inf_{k\geq 0} \{\lambda_k\} > 0$.
Considérons la suite $\{(x^k, p^k)\}_{k=0}^{\infty} \subseteq I\!R^n \times I\!R^m$ définie par

$$\begin{cases} x^k \in \operatorname{argmin}_{x\in I\!R^n} \ \{f_X(x) + \frac{1}{\lambda_k} h^*(\nabla h(p^k) + \lambda_k(Ax-b))\} \\ p^{k+1} = \nabla h^*(\nabla h(p^k) + \lambda_k(Ax^k - b)) \end{cases}$$

et supposons que l'une des deux conditions (H.4.1) ou (H.4.2) est vérifiée.
Sous ces hypothèses, on a les assertions suivantes :

- Si le problème dual (DE) admet au moins une solution alors la suite $\{p^k\}$ converge vers une d'entre elles. De plus, toute valeur d'adhérence de la suite $\{x^k\}$ est une solution du problème (PE).

- Si au contraire le problème (DE) n'admet pas de solution alors la suite $\{p^k\}$ est non bornée.

<u>Preuve du théorème 4.3 :</u>

<u>Pas 1 :</u> Prouvons d'abord la convergence de la suite $\{p^k\}$.

Par la proposition 4.3, nous savons que $p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k)$.
De plus, $\overline{D(\partial(-d))} \subseteq S$ car h est une fonction de Bregman de zone $S = I\!R^m$.

Par conséquent, nous déduisons du théorème 2.7, appliqué à l'opérateur $\partial(-d)$ que la suite $\{p^k\}$ converge vers un zéro de l'opérateur $\partial(-d)$, c'est-à-dire une solution du problème dual (DE) s'il en existe une.
Dans le cas contraire, la suite $\{p^k\}$ est non bornée.

Pas 2 : Supposons à présent que l'ensemble des solutions du problème (DE) est non vide.

Notons $\overline{p}$ la limite de la suite $\{p^k\}$, et $\overline{x}$ une valeur d'adhérence de la suite $\{x^k\}$. (s.p.d.g., nous notons $\{x^k\}$ la sous-suite qui converge vers $\overline{x}$). Nous désirons prouver que $\overline{x}$ est une solution du problème (PE).

a) Dans un premier temps, montrons que $\overline{x}$ est un point admissible pour le problème (PE), c'est-à-dire $A\overline{x} = b$ et $\overline{x} \in X$.

- D'une part, comme la suite $\{x^k\}$ est contenue dans l'ensemble fermé X, il est immédiat que $\overline{x} \in X$.
- D'autre part, il suit de la définition de p^{k+1} que

$$\frac{1}{\lambda_k}(\nabla h(p^k) - \nabla h(p^{k+1})) = \frac{1}{\lambda_k}\{\nabla h(p^k) - \nabla h(\nabla h^*(\nabla h(p^k) + \lambda_k(Ax^k - b)))\} .$$

Or ∇h et ∇h^* sont deux opérateurs inverses (cfr proposition 4.2), ainsi il résulte que

$$\frac{1}{\lambda_k}(\nabla h(p^k) - \nabla h(p^{k+1})) = -(Ax^k - b) . \tag{4.21}$$

De plus, puisque h est continûment différentiable sur $I\!R^m$ et que $\lim_{k\to\infty} p^k = \overline{p}$, nous obtenons que

$$\lim_{k\to\infty} (\nabla h(p^k) - \nabla h(p^{k+1})) = 0 ,$$

et donc par la relation (4.21)

$$\lim_{k\to\infty} -\lambda_k(Ax^k - b) = 0 .$$

Or par hypothèse $\inf_{k\geq 0} \{\lambda_k\} > 0$, d'où $\lim_{k\to\infty} (Ax^k - b) = 0$, c'est-à-dire $A\overline{x} = b$.

b) Dans un second temps, montrons que $\overline{x} \in \text{argmin}_{x\in I\!R^n} \{L(x,\overline{p})\}$ où la fonction lagrangienne $L(x,\overline{p})$ associe à (PE) est définie par $L(x,\overline{p}) = f_X(x) + <\overline{p}, Ax - b>$.

- Montrons d'abord que

$$\overline{x} \in \text{argmin}_{x\in I\!R^n} \{f_X(x) + <\overline{x}, Ax - b>\} \Leftrightarrow -A^T\overline{p} \in \partial f_X(\overline{x}) .$$

En effet, la condition d'optimalité d'un problème de minimisation convexe (cfr proposition A.VII.1), nous assure que

$$\begin{aligned} \overline{x} &\in \operatorname{argmin}_{x \in I\!R^n} \{f_X(x) + < \overline{p}, Ax - b >\} \\ \Leftrightarrow 0 &\in \partial[f_X(x) + < \overline{p}, Ax - b >](\overline{x}) \end{aligned} \tag{4.22}$$

Or la fonction f_1 définie sur $I\!R^n$ par $f_1(x) =< \overline{p}, Ax - b >$ est propre, convexe et $ri(\text{dom } f_1) = I\!R^n$.
Par ailleurs $ri(X) \neq \emptyset$ car X est convexe (cfr proposition A.II.4).
Dès lors il suit que $ri(\text{dom } f_1) \cap ri(\text{dom } f_X) \neq \emptyset$. Ainsi la relation (4.22) est équivalente à

$$0 \in \partial f_X(\overline{x}) + \nabla[< A^T\overline{p}, x >](\overline{x}) \ ,$$

c'est-à-dire

$$0 \in \partial f_X(\bar{x}) + A^T\overline{p} \ ,$$

ou encore

$$-A^T\overline{p} \in \partial f_X(\overline{x}) \ .$$

- Prouvons maintenant que $-A^T\overline{p} \in \partial f_X(\overline{x})$.
D'une part, nous savons par le lemme 4.2 ($1 \Leftrightarrow 2$) que pour tout $k \geq 0$

$$-A^T p^{k+1} \in \partial f_X(x^k)$$

De plus l'opérateur ∂f_X est maximal monotone car f_X est une fonction propre, convexe et s.c.i. (cfr exemple 2.5).
Ainsi nous déduisons du caractère fermé des opérateurs maximaux monotones (cfr proposition 2.4) que $-A^T\overline{p} \in \partial f_X(\overline{x})$.

c) En conclusion, il résulte de a) et b) que $\overline{x}$ vérifie les *conditions de Kuhn Tucker* associées au problème (PE). En effet,

$\overline{x} \in X$
$A\overline{x} = b$
et $\overline{x} \in \operatorname{argmin}_{x \in I\!R^n} \{L(\overline{x}, \overline{p})\}$
où $\overline{p}$ est une solution du dual (DE).

Par conséquent, $\overline{x}$ est une solution du problème (PE). Ce qu'il fallait démontrer.

4.3.5 Cas particulier : $h(x) = \frac{1}{2}||x||^2$

Nous désirons réécrire l'itération de l'algorithme de la méthode des multiplicateurs utilisant la fonction de Bregman h définie de $I\!R^m$ dans $I\!R$ par $h(x) = \frac{1}{2}||x||^2$ dans le cadre d'un problème d'optimisation convexe ne contenant que des contraintes d'égalités.

Il est évident que la zone de la fonction h est $I\!R^m$ et que $R(\nabla h) = I\!R^m$.

Soit $p^0 \in I\!R^m$. Puisque $\nabla h(x) = x$ et que $h^*(y) = \frac{1}{2}||y||^2$ (cfr exemple A.VI.3), l'algorithme (4.5) génère une suite $\{(x^k, p^k)\}_{k=0}^{\infty}$ telle que

$$\begin{cases} x^k \in \text{argmin}_{x \in I\!R^n} \{f_X(x) + \frac{1}{2\lambda_k}||p^k + \lambda_k(Ax - b)||^2\} \\ p^{k+1} = p^k + \lambda_k(Ax^k - b) \end{cases}$$

$$\Leftrightarrow \begin{cases} x^k \in \text{argmin}_{x \in I\!R^n} \{f_X(x) + \frac{1}{2\lambda_k}(||p^k||^2 + \lambda_k^2||Ax - b||^2 + 2\lambda_k < p^k, Ax - b >)\} \\ p^{k+1} = p^k + \lambda_k(Ax^k - b) \end{cases}$$

$$\Leftrightarrow \begin{cases} x^k \in \text{argmin}_{x \in I\!R^n} \{f_X(x) + < p^k, Ax - b > + \frac{1}{2}\lambda_k||Ax - b||^2\} \\ p^{k+1} = p^k + \lambda_k(Ax^k - b) \end{cases}$$

Nous retrouvons ainsi la méthode du *lagrangien augmenté classique* pour le problème (PE) (cfr algorithme 4.1).

4.4 Problème de minimisation convexe avec contraintes d'inégalités

Dans ce paragraphe, nous désirons étudier le problème

$$(PI)\begin{cases} \text{minimiser} & f(x) \\ \text{s.c.} & g_i(x) \leq 0 \quad i = 1, \cdots, m \\ & x \in X \end{cases}$$

où $f, g_i \quad i = 1, \cdots, m$ sont des fonctions propres, convexes et s.c.i., définies de $\mathbb{R}^n$ dans $\mathbb{R} \cup \{+\infty\}$ et X est un ensemble convexe, fermé, non vide de $\mathbb{R}^n$.
Nous supposons également que :

- les fonctions f et $g_i \quad i = 1, \cdots, m$ ont une valeur finie sur X
- l'ensemble dom $f \cap \{x \in X$ tel que $g_i(x) \leq 0 \quad i = 1, \cdots, m\}$ est non vide
- l'hypothèse de qualification de contraintes - il existe $\hat{x} \in X$ tel que $g_i(\hat{x}) < 0$ $i = 1, \cdots, m$ - est vérifiée.

Notre analyse est structurée comme dans le paragraphe précédent.

Nous commençons par énoncer quelques propriétés relatives au dual lagrangien de (PI).

Ensuite nous adaptons l'algorithme de la méthode des multiplicateurs utilisant des fonctions de Bregman au problème (PI).

Enfin, nous vérifions que par le choix des fonctions de Bregman $h_1(p) = \frac{1}{2}||p||^2$ et $h_2(p) = \sum_{i=1}^{m} (p_i \ln p_i - p_i)$, nous retrouvons, respectivement, la méthode du lagrangien augmenté classique pour (PI) (algorithme 4.2), et la méthode exponentielle des multiplicateurs (algorithme 4.3).

4.4.1 Le dual au sens de Lagrange du problème (PI)

Le *problème dual* au sens de Lagrange du problème (PI) s'énonce

$$\text{(DI)} \qquad \begin{cases} \text{Maximiser} & d(p) \\ \text{s.c.} & p \geq 0 \end{cases}$$

où *la fonction duale lagrangienne* associée à (PI) est définie de $I\!R^m$ dans $I\!R \cup \{-\infty\}$ par

$$d(p) = \begin{cases} \inf_{x \in I\!R^n} \; \{L(x,p)\} & \text{si } p \geq 0 \\ -\infty & \text{sinon} \end{cases}$$

c'est-à-dire en vertu de la définition de *fonction lagrangienne* $L(x,p)$ de PI

$$d(p) = \begin{cases} \inf_{x \in I\!R^n} \{f_X(x) + < p, g(x) >\} & \text{si } p \geq 0 \\ -\infty & \text{sinon} \end{cases}$$

Une solution du problème dual (DI) est appelée un *multiplicateur de Lagrange optimal* pour (PI).

Par les mêmes arguments que dans le paragraphe 4.3.1, nous déduisons que la fonction duale associée à (PI) est propre, concave et s.c.s.
Dès lors *le problème* (DI) *est équivalent à la recherche d'un zéro de l'opérateur maximal monotone* $\partial(-d)$.

Etablissons à présent une relation entre les fonctions duale et primale associées au problème (PI).

Définition 4.6 :
La fonction q définie de $I\!R^m$ dans $I\!R \cup \{+\infty\}$ par
$q(u) = \inf_{x \in X} \{f(x) \text{ tel que } g(x) \leq u\}$ est appelée la *fonction primale* associée au problème (PI).

Proposition 4.7 :

$$\forall\, p \in I\!R^m \quad d(p) = -q^*(-p), \text{ c'est-à-dire que } -d = q^* \circ (-I).$$

Preuve : cfr [1, p. 317].

Pour terminer, énonçons une propriété du sous-différentiel de l'opposé de la fonction duale.

Proposition 4.8 :
Soient $\overline{\Omega}^+ := \{p \in I\!R^m \text{ tel que } p \geq 0\}$ et $N_{\overline{\Omega}^+}$: le cône normal à $\overline{\Omega}^+$.
Dès lors $\forall\, p \in I\!R^m \quad \partial(-d)(p) = \partial(-d)(p) + N_{\overline{\Omega}^+}(p)$.

Preuve de la proposition 4.8 :

Soit $p \in I\!R^m$ arbitraire.

- Il est tout d'abord évident que pour tout $y \in \partial(-d)(p)$,

$$y + 0 \in \partial(-d)(p) + N_{\overline{\Omega}^+}(p)\ ,$$

dès lors $\partial(-d)(p) \subseteq \partial(-d)(p) + N_{\overline{\Omega}^+}(p)$.

- Par ailleurs $\partial(-d)(p) + N_{\overline{\Omega}^+}(p) \subseteq \partial(-d)(p)$.
En effet, $\forall\, y \in \partial(-d)(p)$ et $\forall\, u \in N_{\overline{\Omega}^+}(p)$, nous obtenons que

$$-d(z) \geq -d(p) + < y, z - p > \quad \forall\, z \in \text{dom } (-d) \subseteq \overline{\Omega}^+$$

et

$$0 \geq < u, z - p > \quad \forall\, u \in \overline{\Omega}^+\ .$$

Par conséquent

$$-d(z) \geq -d(p) + < y + u, z - p > \quad \forall\, z \in \text{dom } (-d)\ ,$$

c'est-à-dire $y + u \in \partial(-d)(p)$. ∎

4.4.2 Algorithme de la méthode des multiplicateurs utilisant des fonctions de Bregman pour (PI)

Soient h une fonction de Bregman de zone $S \subseteq I\!R^m$ telle que $D(\partial(-d)) \subseteq S$ et $\{\lambda_k\}$ une suite de scalaires strictement positifs.

L'algorithme de la méthode des multiplicateurs utilisant des fonctions de Bregman (algorithme 4.4) génère une suite $\{(x^k, p^k)\}_{k=0}^{\infty} \subseteq I\!R^n \times \overline{\Omega}$ telle que $\forall\ k \geq 0$

$$x^k \in \operatorname{argmin}_{x \in I\!R^n} \ \{L_B(x, p^k, \lambda_k)\} \tag{4.23}$$

$$p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k) \tag{4.24}$$

où le lagrangien augmenté généralisé L_B associé au problème (PI) est défini par

$$L_B(x, p^k, \lambda_k) = \sup_{\xi \geq 0} \{L(x, \xi) - \frac{1}{\lambda_k} D_h(\xi, p^k)\}\ .$$

Avant d'adapter les relations d'itération de la suite $\{(x^k, p^k)\}$ au problème (PI), prenons quelques conventions et énonçons quelques propriétés relatives à la conjuguée monotone d'une fonction de Bregman h.

a) <u>Conventions :</u>

Par la suite, nous désignerons respectivement les fonctions indicatrices des ensembles $\overline{\Omega}^+ := \{p \in I\!R^m \text{ tel que } p \geq 0\}$ et $\overline{\Omega}^- := \{p \in I\!R^m \text{ tel que } p \leq 0\}$ par δ^+ et δ^-, c'est-à-dire $\delta^+ = \delta_{\overline{\Omega}^+}$ et $\delta^- := \delta_{\overline{\Omega}^-}$.

Enonçons de suite une propriété de la conjuguée de Fenchel des fonctions δ^+ et δ^-.

<u>Proposition 4.9 :</u>

1. $(\delta^+)^* = \delta^-$
2. $(\delta^-)^* = \delta^+$

<u>Preuve de la proposition 4.9 :</u>

a) Prouvons que $(\delta^+)^* = \delta^-$

- D'une part considérons $y \in \overline{\Omega}^-$.
Il est immédiat que $\delta^-(y) = 0$. De plus, par la définition de conjuguée de Fenchel (définition 4.1)

$$(\delta^+)^*(y) = \sup_{x \in I\!R^n} \{< x, y > -\delta^+(x)\} ,$$

ou encore

$$(\delta^+)^*(y) = \sup_{x \in \overline{\Omega}^+} \{< x, y >\} .$$

Puisque le suprémum, atteint en $x = 0$, vaut 0,
nous concluons que $(\delta^+)^*(y) = \delta^-(y)$.

- D'autre part, considérons $y \notin \overline{\Omega}^-$.
Il résulte que $\delta^-(y) = +\infty$. Par ailleurs $(\delta^+)^*(y) = \sup_{x \in \overline{\Omega}^-}\{< x, y >\} = +\infty$ car $y \in \overline{\Omega}^+$. D'où $\delta^-(y) = (\delta^+)^*(y)$.

b) La preuve de $(\delta^-)^* = \delta^+$ est analogue à celle de la première égalité. ■

b) Propriétés de la conjuguée monotone d'une fonction de Bregman h

Définition 4.7 : *Conjuguée monotone*
Soit f une fonction convexe définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$. Alors la fonction f^{*+} définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$ par

$$f^{*+}(y) = \sup_{p \geq 0} \{< p, y > -f(y)\}$$

est appelée la *fonction conjuguée monotone* de f.

Proposition 4.10 :
Soit h une fonction de Bregman de zone S telle que $\Omega^+ \subseteq S$. Alors

$$h^{*+}(z) = (h + \delta^+)^*(z) = (h^* \square \delta^-)(z) = \inf_{w \geq z} \{h^*(w)\}$$

où $\square$ désigne *l'opérateur de convolution infimale* (cfr A.VI.6).

Preuve de la proposition 4.10 :

Soit $z \in I\!R^m$ arbitraire

a) Il est évident que $h^{*+}(z) = (h+\delta^+)^*(z)$.

En effet, comme

$$h^{*+}(z) = \sup_{p\geq 0} \{< p, z > -h(p)\}$$

nous avons que

$$h^{*+}(z) = \sup_{p\in I\!R^m} \{< p, z > -(h+\delta^+)p\} ,$$

c'est-à-dire

$$h^{*+}(z) = (h+\delta^+)^* z .$$

b) Montrons maintenant que $(h+\delta^+)^*(z) = (h^*\Box\delta^-)(z)$. Par hypothèse $\Omega^+ \subseteq$dom h, dès lors, il résulte d'une propriété de l'intérieur relatif (proposition A.II.8) (où nous identifions l'ouvert D à Ω^+ et le convexe C à dom h) que

$$\Omega^+ \subseteq ri(\text{dom } h) .$$

D'autre part, il suit de la proposition A.II.5 que $\Omega^+ = ri\overline{\Omega}^+ = ri(\text{dom } \delta^+)$.
Par conséquent, $\Omega^+ \subseteq ri(\text{dom } h) \cap ri(\text{dom } \delta^+)$.
Ainsi, déduisons du théorème relatif à la conjuguée d'une somme [cfr A.VI.7] que $(h+\delta^+)^* = h^*\Box(\delta^+)^*$.
La thèse suit alors de la proposition 4.9.

c) Pour terminer, il suit de la définition de l'opérateur de convolution infimal (cfr A.VI.6) que

$$(h^*\Box\delta^-)z = \inf_{y\in I\!R^m} \{h^*(z-y) + \delta^-(y)\} ,$$

c'est-à-dire

$$(h^*\Box\delta^-)z = \inf_{y\leq 0} h^*(z-y) ,$$

ou encore

$$(h^*\Box\delta^-)z = \inf_{w\geq z} \{h^*(w)\} .$$

■

Définition 4.8 :

Soit f une fonction convexe sur $I\!R^n$. Alors $d \in I\!R^n$ est une *direction de descente* pour f si et seulement si

$$\lim_{\alpha\to\infty} \inf f(x+\alpha d) < +\infty \text{ pour tout } x \in\text{dom } f$$

Proposition 4.11 :

Soit h une fonction de Bregman telle que $S \cap \Omega^+ \neq \emptyset$. Alors la conjuguée de Fenchel de

h, c'est-à-dire h^*, ne possède pas de direction de descente non nulle dans l'orthan positif $\overline{\Omega}^+$.

Preuve de la proposition 4.11 :

Considérons un vecteur $d \in \overline{\Omega}^+ \setminus \{0\}$ arbitraire, et montrons, que pour tout $p \in$dom h^* , $\lim_{\alpha\to\infty}$ inf $h^*(p+\alpha d) = +\infty$.

Soient $u \in \Omega^+ \cap S$ et $z = \nabla h(u)$.

- D'une part puisque S, ∇h et ∇h^* sont des opérateurs inverses (proposition 4.2), il suit que $u = \nabla h^*(z)$.
  De plus, étant donné que h^* est une fonction convexe (proposition 4.1), il résulte que

$$h^*(z+\alpha d) \geq h^*(z) + \alpha d^T \nabla h^*(z)$$

  c'est-à-dire

$$h^*(z+\alpha d) \geq h^*(z) + \alpha d^T u \; . \tag{4.25}$$

- D'autre part $d^T u > 0$ car $u \in \Omega^+ \cap S$ et $d \in \overline{\Omega}^+ \setminus \{0\}$, d'où

$$\lim_{\alpha\to\infty} \text{ inf } \left(h^*(z) + \alpha d^T u\right) = +\infty \; . \tag{4.26}$$

- Finalement en vertu des relations (4.25) et (4.26), nous obtenons qu'il existe $z \in$dom h^* tel que

$$\lim_{\alpha\to\infty} \text{ inf } h^*(z+\alpha d) = +\infty \; . \tag{4.27}$$

  Par ailleurs la fonction h^* est fermée car h est propre et convexe (proposition 4.1). Par conséquent, nous déduisons du théorème A.III.5 que

$$\lim_{\alpha\to\infty} \text{ inf } h^*(p+\alpha d) = +\infty \quad \forall\, p \in \text{dom } h^* \tag{4.28}$$

  En effet, si par l'absurde, il existe $\overline{p} \in$dom h^* tel que

$$\lim_{\alpha\to\infty} \text{ inf } h^*(\overline{p}+\alpha d) < +\infty,$$

  alors il suit du théorème A.III.5 que pour tout $p \in$dom h^*,

$$\lim_{\alpha\to\infty} \text{ inf } h^*(p+\alpha d) < +\infty \; ,$$

  ce qui contredit la relation (4.27).

- En conclusion, le caractère du vecteur $d \in \overline{\Omega}^+ \setminus \{0\}$ et la relation (4.28) nous assurent qu'il n'existe pas de vecteur $d \in \overline{\Omega}^+ \setminus \{0\}$ tel que $\lim_{\alpha\to\infty} \inf \ h^*(p+\alpha d) < \infty$ pour tout $p \in$dom h^*. Ce qu'il fallait démontrer. ■

Proposition 4.12 :

Soit h une fonction de Bregman de zone S telle que $\Omega^+ \subseteq S$.
Alors la fonction h^{*+} est propre, convexe et fermée.
De plus pour tout $z \in$dom h^{*+}, il existe $w \geq z$ tel que $h^{*+}(z) = h^*(w)$.

Preuve de la proposition 4.12 :

a) Remarquons de suite que pour tout $z \in$dom h^{*+} , $h^{*+}(z) = (h^*\square\delta^-)z$ (cfr proposition 4.10).

b) Montrons maintenant que pour tout $d \in I\!R^m \setminus \{0\}$

$$h^* \circ^+ (d) + \delta^- \circ^+ (-d) > 0 \tag{4.29}$$

où $h^*\circ^+$ et $\delta^-\circ^+$ désignent les fonctions de récession des fonctions propres, convexes, fermées h^* et δ^- (cfr définition A.VIII.3).

- D'une part, il résulte de la proposition A.VIII.4 que pour tout $y \in I\!R^m$ et pour tout $x \in \overline{\Omega}^-$

$$\begin{aligned} & (\delta^-\circ^+)y = \lim_{\lambda\to\infty} \frac{\delta^-(x+\lambda y) - \delta^-(x)}{\lambda} \\ \Leftrightarrow \quad & (\delta^-\circ^+)y = \lim_{\lambda\to\infty} \frac{\delta^-(x+\lambda y)}{\lambda} \ . \end{aligned}$$

Dès lors il est évident que

$$(\delta^-\circ^+)y = \begin{cases} 0 & \text{si } y \in \overline{\Omega}^+ \\ +\infty & \text{sinon} \end{cases} \tag{4.30}$$

- D'autre part la proposition 4.11 nous assure qu'il n'existe pas de vecteur $d \in \overline{\Omega}^+ \setminus \{0\}$ tel que

$$\lim_{\alpha\to\infty} \inf \ h^*(x+\alpha d) < +\infty \text{ pour tout } x \in \text{dom } \ h^* \tag{4.31}$$

Ainsi nous déduisons que

$$\forall\, d \in \overline{\Omega}^+ \setminus \{0\} \quad h^* \circ^+ (d) > 0 \ . \tag{4.32}$$

En effet, si par l'absurde il existe $\overline{d} \in \overline{\Omega}^+ \setminus \{0\}$ tel que $h^* \circ^+ (\overline{d}) \leq 0$, il suit du théorème A.VIII.5 que $\lim_{\alpha\to\infty}$ inf $h^*(x + \alpha\overline{d}) < +\infty$ pour tout $x \in$dom h^*, ce qui contredit (4.31).

- Par conséquent il résulte du caractère propre de $h^* \circ^+$ (cfr proposition A.VIII.4) et de (4.30) que pour tout $d \in \overline{\Omega}^+ \setminus \{0\}$ $h^* \circ^+ (d) + \delta^- \circ^+ (-d) = +\infty$.
  De plus il suit des relations (4.30) et (4.32) que pour tout $d \in \overline{\Omega}^+ \setminus \{0\}$ $h^* \circ^+ (d) + \delta^- \circ^+ (-d) = h^* \circ^+ (d) > 0$.
  Ainsi, nous obtenons que $\forall\, d \in I\!R^m \setminus \{0\}$ $h^* \circ^+ (d) + \delta^- \circ^+ (-d) > 0$.

c) Finalement, nous déduisons de (4.29) et de la proposition A.VIII.6 que la fonction $h^{*+} = h^* \Box \delta^-$ est une fonction propre, convexe, fermée et que pour tout $z \in$dom h^{*+}, il existe $\overline{y} \in \overline{\Omega}^-$ tel que

$$h^{*+}(z) = (h^* \Box \delta^-)(z) = \inf_{y \in I\!R^m} \{h^*(z-y) + \delta^-(y)\} = h^*(z - \overline{y}) .$$

Autrement dit, pour tout $z \in$dom h^{*+}, il existe $w \geq z$ tel que $h^{*+}(z) = h^*(w)$. ∎

Proposition 4.13 :

Soit h une fonction de Bregman de zone $S \subseteq I\!R^m$. Alors h^{*+} est une fonction croissante. Plus précisément, si $z \leq z'$ alors $h^{*+}(z) \leq h^{*+}(z')$.

Preuve de la proposition 4.13 :

Soient deux vecteurs z et z' de $I\!R^m$ tels que $z \leq z'$.
Puisque $\{w \in I\!R^m$ tel que $w \geq z\} \subseteq \{w \in I\!R^m$ tel que $w \geq z'\}$,
nous déduisons que $\inf_{w \geq z} \{h^*(w)\} \leq \inf_{w \geq z'} \{h^*(w)\}$.
Ainsi par la proposition 4.10, nous obtenons que $h^{*+}(z) \leq h^{*+}(z')$. ∎

Proposition 4.14 :

Soit h une fonction de Bregman de zone S telle que $\Omega^+ \subseteq S$ et $\Omega^+ \subseteq R(\nabla h)$.
Alors dom $h^{*+} = I\!R^m$.

Preuve de la proposition 4.14 :

- Remarquons d'abord que

$$\text{dom } h^* \subseteq \text{dom } h^{*+} . \tag{4.33}$$

En effet, il suit de la proposition 4.12 que pour tout $z \in$dom h^{*+} il existe $w \geq z$ tel que $w \in$dom h^* et $h^{*+}(z) = h^*(w)$.

- D'autre part, les propositions 4.12 et 4.13, nous assure que la fonction h^{*+} est propre et croissante. Dès lors $\forall\ z \in$dom h^{*+} et $\forall\ u \leq z$, nous obtenons que

$$-\infty \leq h^{*+}(u) \leq h^{*+}(z) \leq +\infty$$

$$\text{c'est-à-dire } \forall\ z \in \text{dom } h^{*+} \text{ et } \forall\ u \leq z, u \in \text{dom } h^{*+}\ . \tag{4.34}$$

- Par conséquent, nous déduisons des relations (4.33) et (4.34) que

$$\text{dom } h^* + \overline{\Omega}^- \subseteq \text{dom } h^{*+}\ . \tag{4.35}$$

En effet, $\forall\ y \in$dom $h^* + \overline{\Omega}^-$, $\exists\ y_1 \in$dom h^* et $y_2 \leq 0$ tel que $y = y_1 + y_2$. Autrement dit, $\forall\ y \in$dom $h^* + \overline{\Omega}^-$, $\exists\ y_1 \in$dom h^* tel que $y \leq y_1$. Dès lors, il suit de (4.33) que $\forall\ y \in$dom $h^* + \overline{\Omega}^-$, $\exists\ y_1 \in$dom h^* tel que $y \leq y_1$, et donc par la relation (4.34), $y \in$dom h^{*+}.

- Par ailleurs il résulte de l'assertion A.IV.9 que $R(\nabla h) \subseteq$dom h^*.
Or par hypothèse $\Omega^+ \subseteq R(\nabla h)$, dès lors

$$\Omega^+ \subseteq \text{dom } h^*\ . \tag{4.36}$$

- Ainsi, regroupant les relations (4.35) et (4.36), nous concluons que

$$\text{dom } h^{*+} \supseteq \text{dom } h^* + \overline{\Omega}^- \supseteq \Omega^+ + \overline{\Omega}^- = I\!R^m\ .$$

Et donc dom $h^{*+} = I\!R^m$.
Ce qu'il fallait démontrer. ■

<u>**Proposition 4.15 :**</u>
Soit h une fonction de Bregman de zone S telle que $\Omega^+ \subseteq S$ et $\Omega^+ \subseteq R(\nabla h)$.
Alors la fonction h^{*+} est différentiable sur $I\!R^m$.

<u>**Preuve de la proposition 4.15 :**</u>

Remarquons dès à présent que $h^{*+} = h^* \Box \delta^-$ (cfr proposition 4.10).

- D'une part vérifions que la fonction h^* est essentiellement régulière (4.37).
En effet, par la définition de fonction de Bregman, h est strictement convexe et différentiable sur S. Dès lors, il suit de la définition 4.3 que la fonction h est essentiellement strictement convexe sur S.
De plus la fonction h étant propre, convexe et continue, elle est fermée [cfr A.III.4].
La thèse suit alors de la proposition 4.4.

- D'autre part, montrons que

$$\Omega^+ \subseteq ri(\text{dom } h^{**}) \cap ri(\text{dom } (\delta^-)^*) \,. \tag{4.38}$$

En effet, il suit de la proposition 4.9 que $(\delta^-)^* = \delta^+$. De plus h étant fermée, il résulte de la proposition A.VI.4 que $h^{**} = h$.
Dès lors

$$ri(\text{dom } h^{**}) \cap ri(\text{dom } (\delta^-)^*) = ri(\text{dom } h) \cap ri(\text{dom } \delta^+) \tag{4.39}$$

Or par hypothèse, $\Omega^+ \subseteq S = \text{dom } h$. Ainsi, nous déduisons d'une propriété de l'intérieur relatif (cfr A.II.8), que $\Omega^+ \subseteq ri(\text{dom } h)$. Par ailleurs, il suit de la proposition A.II.5 que $\Omega^+ = ri\overline{\Omega}^+ = ri(\text{dom } \delta^+)$.
D'où $\Omega^+ \subseteq ri(\text{dom } h) \cap ri(\text{dom } \delta^+)$.
La thèse découle alors de l'égalité (4.39).

- Par conséquent, nous déduisons de (4.37), (4.39) et d'une propriété de la convolution (cfr [1], corollaire 26.3.2) que la fonction $h^{*+} = h^* \square \delta^-$ est essentiellement régulière. Ainsi il résulte des propositions 4.3 et 4.14 que la fonction h^{*+} est différentiable sur $\text{int}(\text{dom } h^{*+}) = I\!R^m$. ■

c) Forme de l'itération

Nous pouvons à présent exprimer les relations d'itération (4.23) et (4.24) de la méthode des multiplicateurs utilisant des fonctions de Bregman en terme des fonctions f_X, h et h^{*+}.

Théorème 4.4 *Forme de l'itération*
Soient $\{\lambda_k\}$ une suite de scalaires strictement positifs, et h une fonction de Bregman de zone S telle que $\overline{\Omega}^+ \subseteq S$ et $\Omega^+ \subseteq R(\nabla h)$.
Considérons la suite $\{(x^k, p^k)\}$ définie par

$$\begin{cases} x^k \in \text{argmin}_{\,x \in I\!R^n} \; \{f_X(x) + \frac{1}{\lambda_k} h^{*+}(\nabla h(p^k) + \lambda_k g(x)\} \\ p^{k+1} = \nabla h^{*+}(\nabla h(p^k) + \lambda_k g(x^k)) \end{cases}$$

Alors

$$\begin{cases} x^k \in \text{argmin}_{\,x \in I\!R^n} \; \{L_B(x, p^k, \lambda_k)\} \\ p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k) \end{cases}$$

où le lagrangien augmenté généralisé L_B associé au problème (PI) est défini par

$$L_B(x, p^k, \lambda_k) = \sup_{\xi \geq 0}\{L(x,\xi) - \frac{1}{\lambda_k} D_h(\xi, p^k)\} .$$

Notons que l'hypothèse $\overline{\Omega}^+ \subseteq S$ assure que $D(\partial(-d)) \subseteq S$. En effet, $D(\partial(-d)) \subseteq \overline{\Omega}^+$.

Dans le but de prouver le théorème 4.4, énonçons les trois lemmes suivants.

<u>**Lemme 4.4 :**</u>
Soient h une fonction de Bregman de zone S telle que $\overline{\Omega}^+ \subseteq S$ et $\Omega^+ \subseteq R(\nabla h)$, p un vecteur de S et λ un scalaire strictement positif.
Considérons l'opérateur P défini sur $I\!R^m$ par

$$P(u) = \frac{1}{\lambda} h^{*+}(\nabla h(p) + \lambda u) .$$

Alors

1. L'opérateur P, de domaine $I\!R^m$, est propre, convexe, fermé et croissant. De plus P est différentiable et
$$\nabla P(u) = \nabla h^{*+}(\nabla h(p) + \lambda u) .$$

2. Pour tout $v \in S$, la fonction conjuguée de Fenchel de P est définie par
$$P^*(v) = \frac{1}{\lambda}[h(v) + \delta^+(v) - v^T \nabla h(p)] .$$
De plus
$$\partial P^*(v) = \frac{1}{\lambda}\{\nabla h(v) - \nabla h(p)\} + N_{\overline{\Omega}^+}(v) ,$$
où $N_{\overline{\Omega}^+}$ désigne le cône normal à $\overline{\Omega}^+$.

<u>**Preuve du lemme 4.4 :**</u>

a) L'assertion (1) résulte immédiatement des propositions 4.12, 4.13, 4.14, 4.15.

b) Soit un vecteur $v \in S$ arbitraire. Prouvons que $P^*(v) = \frac{1}{\lambda}[h(v) + \delta^+(v) - v^T \nabla h(p)]$.
Il suit de la définition de la conjuguée de Fenchel (définition 4.1) que

$$P^*(v) = \sup_{u \in I\!R^m} \{v^T u - \frac{1}{\lambda} h^{*+}(\nabla h(p) + \lambda u)\} .$$

Définissant alors $w = \nabla h(p) + \lambda u$, nous obtenons que

$$P^*(v) = \frac{1}{\lambda} \sup_{w \in I\!R^m} \{v^T(w - \nabla h(p)) - h^{*+}(w)\} ,$$

dès lors

$$P^*(v) = \frac{1}{\lambda}(h^{*+})(v) - \frac{1}{\lambda} v^T \nabla h(p) .$$

Or par la proposition 4.10, $h^{*+} = h^* \Box \delta^-$, par conséquent il résulte que

$$P^*(v) = \frac{1}{\lambda}[(h^* \Box \delta^-)^*(v) - v^T \nabla h(p)] .$$

De plus, nous déduisons du théorème relatif à la conjuguée d'une convolution (cfr A.VI.7) que

$$P^*(v) = \frac{1}{\lambda}[h^{**}(v) + (\delta^-)^*(v) - v^T \nabla h(p)] .$$

Or $h^{**} = h$ car h est fermée (cfr proposition A.VI.4) et $(\delta^-)^* = \delta^+$ (cfr proposition 4.9), finalement, nous concluons que

$$P^*(v) = \frac{1}{\lambda}[h(v) + \delta^+(v) - v^T \nabla h(p)] .$$

c) Pour terminer, montrons que

$$\partial P^*(v) = \frac{1}{\lambda}\{\nabla h(v) - \nabla h(p)\} + N_{\overline{\Omega}^+}(v) \quad \forall\, v \in S$$

Soit un vecteur $v \in S$ arbitraire. Par le point b), nous savons que

$$\partial P^*(v) = \partial[\frac{1}{\lambda}(h(v) + \delta^+(v) - v^T \nabla h(p))] .$$

Or par hypothèse $\Omega^+ \subseteq \overline{\Omega}^+ \subseteq S$ =dom h, ainsi, nous déduisons d'une propriété de l'intérieur relatif (cfr A.II.8) que $\Omega^+ \subseteq ri(\text{dom } h)$. D'autre part $\Omega^+ = ri\overline{\Omega}^+ = ri$ dom δ^+ (cfr proposition A.II.5). Par conséquent nous obtenons que $\Omega^+ \subseteq ri(\text{dom } h) \cap ri(\text{dom } \delta^+)$.

Le théorème relatif au sous-différentiel d'une somme (cfr A.IV.12) nous assure donc que

$$\partial P^*(v) = \frac{1}{\lambda}\{\nabla h(v) + \partial \delta^+(v) - \nabla h(p)\} .$$

Or $\partial \delta^+ = N_{\overline{\Omega}^+}$ (cfr A.V.2), de plus par la définition d'un cône $\frac{1}{\lambda} N_{\overline{\Omega}^+} = N_{\overline{\Omega}^+}$. Finalement, nous concluons que

$$\partial P^*(v) = \frac{1}{\lambda}\{\nabla h(v) - \nabla h(p)\} + N_{\overline{\Omega}^+}(v) .$$

■

Lemme 4.5 :
Soient h une fonction de Bregman de zone S telle que $\overline{\Omega}^+ \subseteq S$ et $\Omega^+ \subseteq R(\nabla h)$, p un vecteur de S et λ un scalaire strictement positif.
Considérons l'opérateur P défini sur $I\!R^m$ par

$$P(u) = \frac{1}{\lambda} h^{*+}(\nabla h(p) + \lambda u) \ .$$

Alors

$$\inf_{x \in X} \ \{f(x) + P(g(x))\} = \inf_{u \in I\!R^m} \ \{q(u) + P(u)\}$$

où q est la fonction primale associée au problème (PI) (cfr définition 4.6).

La preuve du résultat se base sur la caractérisation de la borne inférieure d'un ensemble, rappelons donc cette propriété.

Proposition 4.16 :
Soit $S \subseteq I\!R$ alors

$$m = \inf \ S \Leftrightarrow \ \forall \, \varepsilon > 0 \quad [m, m+\varepsilon[\cap S \neq \emptyset$$

Preuve du lemme 4.5 :

Pas 1 : Dans un premier temps, prouvons que

$$\inf_{x \in X} \ \{f(x) + P(g(x))\} = \inf_{u \in I\!R^m} \ G(u)$$
$$\text{où } G(u) = \inf_{x \in X} \ \{f(x) + P(g(x)) \text{ tel que } g(x) \leq u\} \ ; \tag{4.40}$$

a) Puisque $\forall \, u \in I\!R^m \{x \in X \text{ tel que } g(x) \leq u\} \subseteq X$, il est immédiat que

$$\begin{aligned} \forall \, u \in I\!R^m \quad \inf_{x \in X} \{f(x) + P(g(x))\} &\leq \inf_{x \in X} \{f(x) + P(g(x)) \\ &\text{tel que } g(x) \leq u\} = G(u) \end{aligned} \tag{4.41}$$

b) Posons $m := \inf_{x \in X} \ \{f(x) + P(g(x))\}$.
Il suit de la caractérisation de la borne inférieure (proposition 4.16) que

$$\forall \, \varepsilon > 0 \quad \exists \quad \hat{x} \in X \text{ tel que } m \leq f(\hat{x}) + P(g(\hat{x})) < m + \varepsilon \tag{4.42}$$

Posant $\hat{u} = g(\hat{x})$, nous obtenons que

$$G(\hat{u}) \leq f(\hat{x}) + P(\hat{u}) < m + \varepsilon \ .$$

Or par la relation (4.42), $G(\hat{u}) \geq m$.
Ainsi, il résulte que

$$\forall\, \varepsilon > 0 \;\exists\quad \hat{u} \in I\!R^m \text{ tel que } m \leq G(\hat{u}) < m + \varepsilon\ .$$

La thèse suit alors de la condition suffisante de la proposition 4.16.

Pas 2 : Montrons à présent que

$$\inf{}_{u \in I\!R^m}\, G(u) = \inf{}_{v \in I\!R^m}\, \{q(v) + P(v)\}\ .$$

a) Remarquons que

$$\inf{}_{u \in I\!R^m}\, G(u) \leq q(v) + P(v)\ \forall\, v \in I\!R^m\ . \tag{4.43}$$

En effet, soit v un vecteur arbitraire de $I\!R^m$.
Il suit de la définition de la fonction G (cfr (4.40)) et du caractère croissant de l'opérateur P (cfr lemme 4.4) que

$$G(v) \leq \inf{}_{x \in X}\ \{f(x) + P(v) \text{ tel que } g(x) \leq v\}\ ,$$

c'est-à-dire

$$G(v) \leq \inf{}_{x \in X}\ \{f(x) \text{ tel que } g(x) \leq v\} + P(v)\ ,$$

Ainsi par la définition de la fonction q, nous obtenons que $G(v) \leq q(v) + P(v)$, d'où $\inf{}_{u \in I\!R^m}\, G(u) \leq q(v) + P(v)\ \forall\, v \in I\!R^m$.

b) Notons $m := \inf{}_{u \in I\!R^m}\, G(u)$ et vérifions que

$$\forall\, \varepsilon > 0\ \ [m, m + \varepsilon[\cap \{q(v) + P(v) \text{ tel que } v \in I\!R^m\} \neq \emptyset \tag{4.44}$$

- D'une part $m = \inf{}_{u \in I\!R^m}\, G(u)$, d'où

$$\forall\, \varepsilon > 0 \text{ il existe } \overline{u} \in I\!R^m \text{ tel que } m \leq G(\overline{u}) < m + \varepsilon \tag{4.45}$$

- D'autre part il suit de la définition de $G(\overline{u})$ (cfr relation (4.40)) que

$$\forall\, \varepsilon' > 0\quad \exists\quad \hat{x} \in X \text{ tel que } G(\overline{u}) \leq f(\hat{x}) + P(g(\hat{x})) < G(\overline{u}) + \varepsilon' \tag{4.46}$$

- Finalement, en considérant $\varepsilon' = \varepsilon$ dans (4.45) et (4.46), nous obtenons que

  $$\forall\, \varepsilon > 0\ \ \exists\quad \overline{u} \in I\!R^m \text{ et } \exists\quad \hat{x} B X \text{ tel que } f(\hat{x}) + P(g(\hat{x})) < G(\overline{u}) + \varepsilon < m + 2\varepsilon\ ,$$

  d'où en posant $\hat{v} = g(\hat{x})$,

$$\forall\, \varepsilon > 0\ \exists\ \hat{v} \in I\!R^m \text{ tel que } f(\hat{x}) + P(\hat{v}) < m + 2\varepsilon \tag{4.47}$$

Par ailleurs, il suit de la définition de l'infimum que

$$\inf_{x \in X} \; \{f(x) \text{ tel que } g(x) \leq \hat{v}\} + P(\hat{v}) \leq f(\hat{x}) + P(\hat{v}) \tag{4.48}$$

Par conséquent, nous déduisons de (4.47), (4.48) et de la définition de la fonction primal q que $\forall\, \varepsilon > 0 \;\; \exists \;\; \hat{v} \in I\!R^m$ tel que $q(\hat{v}) + P(\hat{v}) < m + 2\varepsilon$. De plus par la relation (4.43),

$$q(\hat{v}) + P(\hat{v}) \geq \inf_{u \in I\!R^m} G(u) := m$$

En conclusion, $\forall\, \varepsilon > 0 \; [m, m + \varepsilon[\cap \{q(v) + P(v) \text{ tel que } v \in I\!R^m\} \neq \emptyset$.
La thèse suit alors de la caractérisation de la borne inférieure d'un ensemble (cfr proposition 4.16). ■

Lemme 4.6 :
Soient $\{\lambda_k\}$ une suite de scalaires strictement positifs et h une fonction de Bregman de zone S telle que $\overline{\Omega}^+ \subseteq S$ et $\Omega^+ \subseteq R(\nabla h)$.
Considérons la suite $\{(x^k, p^k)\}_{k=0}^{\infty}$ définie par

$$\begin{cases} x^k \in \operatorname{argmin}_{x \in I\!R^n} \{fx(x) + \frac{1}{\lambda_k} h^{*+}(\nabla h(p^k) + \lambda_k g(x))\} \\ p^{k+1} = \nabla h^{*+}(\nabla h(p^k) + \lambda_k g(x^k)) \end{cases}$$

Alors

$$g(x^k) - \frac{1}{\lambda_k}[\nabla h(p^{k+1}) - \nabla h(p^k)] \in N_{\overline{\Omega}^+}(p^{k+1}) \quad \forall\, k \geq 0$$

Preuve du lemme 4.6 :

Considérons l'opérateur P_k défini sur $I\!R^m$ par

$$P_k(u) = \frac{1}{\lambda_k} h^{*+}(\nabla h(p^k) + \lambda_k u) \quad \forall\, k \geq 0$$

- D'une part, remarquons que

$$p^k \in S \quad \forall\, k \geq 0 \tag{4.49}$$

  En effet, en exprimant la relation d'itération de p^k en terme de l'opérateur P_k, nous obtenons que $p^{k+1} = \nabla P_k(g(x^k))$.
  Or par le lemme 4.4 P_k est croissant, dès lors $p^{k+1} \geq 0$, ie $p^{k+1} \in \overline{\Omega}^+$.
  La thèse suit alors de l'hypothèse $\overline{\Omega}^+ \subseteq S$.

- D'autre part, il résulte du lemme 4.4 que l'opérateur P_k est propre, convexe, fermé et différentiable.

Par conséquent ∇P_k et $\partial(P_k)^*$ sont des opérateurs inverses (cfr proposition 4.2), et donc

$$p^{k+1} = \nabla P_k(g(x^k)) \Leftrightarrow g(x^k) \in \partial(P_k)^*(p^{k+1}) .$$

Or par la relation (4.49) $p^{k+1} \in S$,
ainsi nous déduisons du lemme 4.4 que

$$g(x^k) \in \frac{1}{\lambda_k}[\nabla h(p^{k+1}) - \nabla h(p^k)] + N_{\overline{\Omega}^+}(p^{k+1}) .$$

■

Nous disposons maintenant de tous les résultats nécessaires pour prouver le théorème relatif à la forme de l'itération de la méthode des multiplicateurs avec fonctions de Bregman pour le problème (PI).

preuve du théorème 4.4 :

Pas 1 : Prouvons d'abord que

$$x^k \in \text{argmin}_{x\in I\!R^n} \{L_B(x,p^k,\lambda_k)\} \Leftrightarrow x^k \in \text{argmin}_{x\in I\!R^n} \{f_X(x)+\frac{1}{\lambda_k}h^{*+}(\nabla h(p^k)+\lambda_k g(x))\} .$$

Comme,

$$L_B(x,p^k,\lambda_k) = \sup_{\xi\geq 0} \{L(x,\xi) - \frac{1}{\lambda_k}D_h(\xi,p^k)\} ,$$

il suit des définitions du lagrangien associé à (PI) et de la fonction D_h que

$$\begin{aligned} L_B(x,p^k,\lambda_k) &= \sup_{\xi\geq 0} \{f_X(x)+ < \xi, g(x) > -\frac{1}{\lambda_k}h(\xi) + \frac{1}{\lambda_k}h(p^k) \\ &+ \frac{1}{\lambda_k} < \nabla h(p^k), \xi - p^k >\} \end{aligned}$$

c'est-à-dire

$$\begin{aligned} L_B(x,p^k,\lambda_k) &= f_X(x) + \frac{1}{\lambda_k}h(p^k) - \frac{1}{\lambda_k} < \nabla h(p^k), p^k > \\ &+ \frac{1}{\lambda_k}\sup_{\xi\geq 0} \{< \nabla h(p^k) + \lambda_k g(x), \xi > -h(\xi)\} . \end{aligned}$$

ou encore

$$\begin{aligned} L_B(x,p^k,\lambda_k) &= f_X(x) + \frac{1}{\lambda_k}h(p^k) - \frac{1}{\lambda_k} < \nabla h(p^k), p^k > \\ &+ \frac{1}{\lambda_k}h^{*+}(\nabla h(p^k) + \lambda_k g(x)) \end{aligned}$$

La thèse suit alors de l'indépendance en x du terme $\frac{1}{\lambda_k}h(p^k) - \frac{1}{\lambda_k} < \nabla h(p^k), p^k >$.

Pas 2 : Montrons à présent que

$$\text{si } p^{k+1} = \nabla h^{*+}(\nabla h(p^k) + \lambda_k g(x^k)) \text{ ,alors } p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k) \;.$$

a) Considérons l'opérateur P_k défini sur $I\!R^m$ par $P_k(u) = \frac{1}{\lambda_k} h^{*+}(\nabla h(p^k) + \lambda_k u)$ $\forall\; k \geq 0$. Il est évident que

$$x^k \in \text{argmin}_{\;x \in I\!R^n} \{f_X(x) + P_k(g(x))\} \tag{4.50}$$

$$p^{k+1} = \nabla P_k(g(x^k)) \tag{4.51}$$

b) Définissons maintenant $u^k = g(x^k)$ $\forall\; k \geq 0$ et montrons que

$$-\,u^k \in \partial(-d)(p^{k+1}) \tag{4.52}$$

- D'une part, il résulte du lemme (4.5) que

  $$\inf_{\;x \in I\!R^n} \{f_X(x) + P_k(g(x))\} = \inf_{\;u \in I\!R^m} \{q(u) + P_k(u)\} \;.$$

  Ainsi, puisque par la relation (4.50), x^k minimise $f(x) + P_k(g(x))$ sur X, il est immédiat que $u^k = g(x^k)$ minimise $q(u) + P_k(u)$ sur $I\!R^m$.
  Par conséquent

  $$0 \in \partial(q + P_k)(u^k) \tag{4.53}$$

- D'autre part $ri(\text{dom}\;\; q) \cap ri(\text{dom}\;\; P_k) \neq \emptyset$. En effet, $ri(\text{dom}\;\; q) \subseteq I\!R^m$ est non vide car dom q est convexe (cfr proposition A.II.4).
  De plus, il suit du lemme 4.4 que dom $(P_k) = I\!R^m$.
  Ainsi la relation (4.53) et la différentiabilité de P_k (cfr lemme 4.4) nous assurent que

  $$0 \in \partial q(u^k) + \nabla P_k(u^k) \;,$$

  c'est-à-dire

  $$-\nabla P_k(u^k) \in \partial q(u^k)$$

  ou encore

  $$-\nabla P_k(g(x^k)) \in \partial q(u^k)$$

  et donc

  $$-p^{k+1} \in \partial q(u^k)$$

  De plus, puisque ∂q et ∂q^* sont des opérateurs inverses [proposition 4.2], nous obtenons que

  $$u^k \in \partial q^*(-p^{k+1}) \tag{4.54}$$

- Pour terminer, remarquons que

$$\partial(-d) = (-I) \circ \partial q^* \circ (-I) \quad \text{où } I \text{ désigne l'identité} \tag{4.55}$$

En effet, il suit de la proposition 4.7 que $-d = q^* \circ (-I)$.
Dès lors $\partial(-d) = \partial(q^* \circ (-I))$.
Or $(R(I) \cap ri(\text{dom} \quad q^*)) = ri(\text{dom} \quad q^*)$ est non vide car q^* est convexe [cfr proposition A.II.4).
Ainsi, il résulte de la proposition A.IV.13 que
$\partial(-d) = \partial(q^* \circ (-I)) = (-I) \circ \partial q^* \circ (-I)$.

- Par conséquent, les relations (4.54) et (4.55), nous assurent que

$$u^k \in \partial q^*(-p^{k+1}) \Leftrightarrow -u \in \partial(-d)(p^{k+1}) \, .$$

c) Montrons à présent que $p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k)$ pour tout $k \geq 0$.
Définissons $w^{k+1} = u^k - \frac{1}{\lambda_k}(\nabla h(p^{k+1}) - \nabla h(p^k)) \; \forall \; k \geq 0$.
Dès lors

$$-u^k = -\{\frac{1}{\lambda_k}(\nabla h(p^{k+1}) - \nabla h(p^k)) + w^{k+1}\} \, .$$

Or par (4.52) nous savons que $-u^k \in \partial(-d)(p^{k+1})$, ainsi, il suit que

$$\frac{1}{\lambda_k}(\nabla h(p^k) - \nabla h(p^{k+1})) \in \partial(-d)(p^{k+1}) + w^{k+1} \, .$$

Par ailleurs, il résulte du lemme 4.6 que $w^{k+1} \in N_{\overline{\Omega}^+}(p^{k+1})$.
Par conséquent nous obtenons que

$$\frac{1}{\lambda_k}[\nabla h(p^k) - \nabla h(p^{k+1})] \in \partial(-d)(p^{k+1}) + N_{\overline{\Omega}^+}(p^{k+1}) \, .$$

De plus, la proposition 4.8 nous assure que $\forall \; p$

$$\partial(-d)(p) = \partial(-d)(p) + N_{\overline{\Omega}^+}(p) \, ,$$

d'où

$$\frac{1}{\lambda_k}(\nabla h(p^k) - \nabla h(p^{k+1})) \in \partial(-d)(p^{k+1}) \, ,$$

ce qui se réécrit

$$p^{k+1} \in ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k) \, .$$

Or l'opérateur $((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)$ est univoque car l'opérateur $\partial(-d)$ est monotone (cfr proposition 2.7). Ainsi, nous concluons que

$$p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k) \, .$$

Ce qu'il fallait démontrer. ∎

Il est maintenant temps d'énoncer l'algorithme de *la méthode des multiplicateurs utilisant des fonctions de Bregman pour un problème avec contraintes d'inégalités.*

<u>Algorithme 4.6 :</u>

Soient $\{\lambda_k\}_{k=0}^{\infty}$ une suite de scalaires strictement positifs, h une fonction de Bregman telle que $\overline{\Omega}^+ \subseteq S$ et $\Omega^+ \subseteq R(\nabla h)$.
A l'itération k, soient $p^k \in \overline{\Omega}^+$ et $\lambda_k > 0$ connus.

1. Résoudre le problème de minimisation

$$x^k \in \text{argmin}_{\, x\in \mathbb{R}^n} \left\{ f_X(x) + \frac{1}{\lambda_k} h^{*+}(\nabla h(p^k) + \lambda_k g(x)) \right\} .$$

2. Calculer l'itération des multiplicateurs

$$p^{k+1} = \nabla h^{*+}(\nabla h(p^k) + \lambda_k g(x^k)) .$$

4.4.3 Conditions d'existence et d'unicité des itérés

En vertu du théorème 4.4, nous savons que la relation d'itération de la suite des multiplicateurs peut se réécrire $p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k)$.
L'existence de la suite $\{p^k\}$ est donc assurée par des conditions semblables à celles du théorème 2.6, en l'occurrence :

$$\begin{aligned} & R(\nabla h) = \mathbb{R}^m \\ \text{ou,}\quad & R(\nabla h) \text{ est un ouvert et } 0 \in R(\partial(-d)) . \end{aligned}$$

Par contre, nous ne connaissons pas de conditions d'existence de la suite $\{x^k\}$ dans le cas général. Toutefois, lorsque la fonction f est *fortement convexe*, alors la fonction définie sur $\mathbb{R}^n$ par $f(x) + \frac{1}{\lambda_k} h^{*+}(\nabla h(p^k) + \lambda_k g(x))$ est elle aussi fortement convexe. Ainsi, il suit de la proposition 3.2 qu'il existe un seul $x \in X$ tel que

$$x^k = \text{argmin}_{\, x\in X} \left\{ f(x) + \frac{1}{\lambda_k} h^{*+}(\nabla h(p^k) + g(x)) \right\} .$$

4.4.4 Convergence vers une solution des problèmes primal (PI) et dual (DI)

Enonçons directement le théorème de convergence de la méthode des multiplicateurs utilisant des fonctions de Bregman vers une solution des problèmes (PI) et (DI).

Théorème 4.5 :
Soient $\{\lambda_k\}_{k=0}^{\infty}$ une suite de scalaires strictement positifs telle que $\inf_{k\geq 0} \{\lambda_k\} > 0$ et h une fonction de Bregman de zone S telle que $\overline{\Omega}^+ \subseteq S$ et $\Omega^+ \subseteq R(\nabla h)$.
Supposons que la suite $\{(x^k, p^k)\}_{k=0}^{\infty} \subseteq I\!R^n \times \overline{\Omega}^+$

$$\begin{cases} x^k \in \operatorname{argmin}_{x \in I\!R^n} \{f_X(x) + \frac{1}{\lambda_k} h^{*+}(\nabla h(p^k) + \lambda_k g(x))\} \\ p^{k+1} = \nabla h^{*+}(\nabla h(p^k) + \lambda_k g(x^k)) \end{cases}$$

est bien définie.
Sous ces hypothèses, on a les assertions suivantes :

- Si le problème dual (DI) admet au moins une solution alors la suite $\{p^k\}$ converge vers une d'entre elles.
  De plus si les fonctions g_i $i = 1, \cdots, m$ sont continues sur X, alors toute valeur d'adhérence de la suite $\{x^k\}$ est solution du problème (PI).

- Si au contraire le problème (DI) n'admet pas de solution alors la suite $\{p^k\}$ est non bornée.

Dans le but de prouver ce théorème, nous énonçons une propriété du sous-gradient de la composée de fonctions convexes.

Proposition 4.17 :
Soient $f_1, \cdots, f_m$ des fonctions propres et convexes sur $I\!R^n$ telles que $\cap_{i=1}^m ri(\text{dom } f_i) \neq \emptyset$.
Notons $f(x) = (f_1(x), \cdots, f_m(x)) \in (-\infty, +\infty]^m \ \forall\, x$.
Soit f_0 une fonction concave et croissante, c'est-à-dire que $u \leq v \Rightarrow f_0(u) \leq f_0(u)$, et différentiable sur $I\!R^m$.
Considérons un vecteur $x^0 \in \cap_{i=1}^m (\text{dom } f_i)$.

Définissons une fonction G de $I\!R^n$ dans $I\!R \cup \{+\infty\}$ par

$$G(x) = \begin{cases} f_0(f(x)) & \text{si } f(x) \in I\!R^m \\ +\infty & \text{si } f_i(x) = +\infty \quad \forall\, i = 1, \cdots, m \end{cases}$$

et une fonction L définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$ par $L(x) = \sum_{i=1}^m y_i f_i(x)$ où $y = \nabla f_0(f(x^0))$.
De plus nous prenons comme convention que $y_i f_i(x) = +\infty$ si $y_i = 0$ et $f_i(x) = +\infty$.
Alors

1. La fonction G est propre et convexe.
2. $\forall\, d \in I\!R^n$ $L'(x^0; d) = G'(x^0; d)$ où $L'(x^0; d)$ désigne la dérivée directionnelle de L en x^0 dans la direction d.
3. $\partial G(x^0) = \sum_{i=1}^m [\nabla f_0(f(x^0))]_i \partial f_i(x^0) + N_{(\mathrm{dom}\ G)}(x^0)$.

Preuve de la proposition 4.16 :

Pas 1 : [La fonction G est propre et convexe].

- Soient u et v deux vecteurs de $\cap_{i=1}^m(\mathrm{dom}\ \ f_i)$ tels que $u \neq v$ et $\lambda \in [0, 1]$.
  Puisque $\forall\, i = 1, \cdots, m$ f_i est une fonction convexe, nous obtenons que

  $$f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda) f(v)\ .$$

  Ainsi, il suit de la croissance et de la convexité de f_0 que

  $$f_0(f(\lambda u + (1-\lambda)v)) \leq f_0(\lambda f(u) + (1-\lambda) f(v)) \leq \lambda f_0(f(u)) + (1-\lambda) f_0(f(v))\ .$$

  D'où,

  $$G(\lambda u + (1-\lambda)v) \leq \lambda G(u) + (1-\lambda) G(v)\ .$$

  Ce qui prouve la convexité de G.

- D'autre part, G est propre.
  En effet, il est évident que $G(x) > -\infty$ $\forall\, x$. De plus par hypothèse dom $\ f_0 = I\!R^m$ et il existe $x \in I\!R^n$ tel que $f(x) \in I\!R^m$.

<u>Pas 2 :</u> $[G'(x^0;d) = L'(x^0;d)$ pour tout $d \in I\!R^n]$

a) Remarquons dès à présent que $\forall\, d \in I\!R^n$ et $\forall\, t \in I\!R$

$$\begin{aligned} L(x^0+td) \in I\!R \;&\Leftrightarrow\; f_i(x^0+td) \in I\!R \quad \forall\, i = 1,\cdots,m \\ &\Leftrightarrow\; f_0(f(x^0+td)) \in I\!R \\ &\Leftrightarrow\; G(x^0+td) \in I\!R \end{aligned} \tag{4.56}$$

En effet, dom $f_0 = I\!R^m$ car f_0 est différentiable.

b) Supposons d'abord que $\forall\, t > 0\ \exists\ t_0 < t$ tel que $L(x^0+t_0d) = +\infty$.
Nous pouvons alors construire une suite $(t_k)_{k=0}^{\infty} \subseteq I\!R$ telle que

$$\lim_{k\to\infty} t_k = 0 \text{ et } L(x^0+t_kd) = +\infty \text{ pour tout } k \geq 0 \tag{4.57}$$

Or $L(x^0) \in I\!R$ car par définition de $x^0, f(x^0) \in I\!R^m$.
D'où $\lim_{t_k\to 0} \frac{L(x^0+t_kd)-L(x^0)}{t_k} = +\infty$.
Par ailleurs en vertu de (4.56) et (4.57), on a $G(x^0+t_kd) = +\infty$ pour tout $k \geq 0$.
Or $G(x^0) \in I\!R$ puisque $f(x^0) \in I\!R^m$ et dom $f^0 = I\!R^m$, dès lors

$$\lim_{t_k\to 0} \frac{G(x^0+t_kd)-G(x^0)}{t_k} = +\infty\,.$$

Par conséquent, il résulte que $L'(x^0;d) = G'(x^0;d) = +\infty$.

c) Supposons maintenant que $\exists\ t_0$ tel que $\forall\, t \leq t_0\ L(x^0+td) \neq +\infty$.

- Calculons d'abord $G'(x^0;d)$. Par définition de G on a

$$\lim_{t\to 0} \frac{G(x^0+td)-G(x^0)}{t} = \lim_{t\to 0} \frac{f_0(f(x^0+td))-f_0(f(x^0))}{t}\,.$$

Or f_0 est différentiable sur $[f(x^0), f(x^0+td)] \subseteq I\!R^m$.
Ainsi, par le théorème des accroissements finis et la continuité du produit scalaire, nous obtenons que

$$\lim_{t\to 0} \frac{G(x^0+td)-G(x^0)}{t} =< \lim_{t\to 0} \nabla f_0(c_t), \lim_{t\to 0} \frac{f(x^0+td)-f(x^0)}{t} > \tag{4.58}$$

où $c_t \in]f(x^0), f(x^0+td)[$.
De plus,

$$\lim_{t\to 0} \nabla f_0(c_t) = \nabla f_0(f(x^0)) \tag{4.59}$$

En effet, puisque f_0 est convexe et différentiable sur $I\!R^m$, il suit de [19, théorème 25.5] que ∇f_0 est continu. Or $\lim_{t\to 0} c_t = f(x^0)$ car $\lim_{t\to 0} f(x^0+td) = f(x_0)$ (cfr [19], théorème 10.1).
Il découle donc des relations (4.58) et (4.59) que

$$G'(x^0;d) = \lim_{t\to 0} \frac{G(x^0+td)-G(x^0)}{t} =< \nabla f_0(f(x^0)), \lim_{t\to 0} \frac{f(x^0+td)-f(x^0)}{t} > \tag{4.60}$$

- Calculons ensuite $L'(x^0;d)$. Par définition de L, on a

$$L(x^0+td) - L(x^0) = \sum_{i=1}^{m} y_i(f_i(x^0+td) - f_i(x^0)) .$$

Or $y = \nabla f_0(f(x^0))$, d'où $L(x^0+td)-L(x^0) =< \nabla f_0(f(x^0)), f(x^0+td)-f(x^0) >$.
Ainsi, il suit de la continuité du produit scalaire que

$$L'(x^0;d) = \lim_{t\to 0} \frac{L(x^0+td)-L(x^0)}{t} =< \nabla f_0(f(x^0)), \lim_{t\to 0} \frac{f(x^0+td)-f(x^0)}{t} > \tag{4.61}$$

Par conséquent, il résulte des relations (4.60), (4.61) que $L'(x^0;d) = G'(x^0;d)$.

<u>Pas 3 :</u> $[\partial G(x^0) = \sum_{i=1}^m y_i \partial f_i(x^0) + N_{(\text{dom } G)}(x^0)]$.

Remarquons dès à présent que

$$y = \nabla f_0(f(x^0)) \text{ est positif car } f_0 \text{ est croissante.} \tag{4.62}$$

Soit l'ensemble $Q := \{z \text{ tel que } z \in \sum_{i=1}^m y_i \partial f_i(x^0) + N_{(\text{dom } G)}(x^0)\}$.

a) Prouvons tout d'abord que $Q \subseteq \partial G(x^0)$.
Soient $S^i \in \partial f_i(x^0)$ $i = 1,\cdots,m$ et S la matrice dont les lignes sont les vecteurs s^i.
En vertu de la définition du sous-différentiel, nous obtenons que

$$\forall\, x \in I\!R^n \quad f_i(x) \geq f_i(x^0) + < s^i, x - x^0 > \quad i = 1,\cdots,m$$

autrement dit,

$$\forall\, x \in I\!R^n \; f(x) \geq f(x^0) + S(x-x^0)$$

Or la fonction f_0 est croissante et différentiable sur $I\!R^m$, d'où

$$\forall\, x \in I\!R^n f_0(f(x)) \geq f_0(f(x^0) + S(x-x^0)) \geq f_0(f(x)) + [\nabla f_0(f(x^0))]^T S(x-x^0) ,$$

c'est-à-dire

$$\forall\ x \in I\!R^n\ G(x) \geq G(x^0) + y^T S(x - x^0)\ ,$$

et donc $y^T S \in \partial G(x^0)$.
Il suit alors du caractère arbitraire des vecteurs $s^i \in \partial f_i(x^0)$ $i = 1, \cdots, m$ que

$$\sum_{i=1}^{m} y_i \partial f_i(x^0) \subseteq \partial G(x^0)\ .$$

Ainsi, puisque $\partial G(x^0) = \partial G(x^0) + N_{\text{dom } G}(x^0)$ (cfr proposition A.IV.14) nous obtenons que

$$Q := \sum_{i=1}^{m} y_i \partial f_i(x^0) + N_{\text{dom } G} \subseteq \partial G(x^0)\ .$$

b) Montrons ensuite par *contraposition* que $\partial G(x^0) \subseteq Q$, c'est-à-dire, vérifions que $\forall\ w \in I\!R^n \setminus Q, w \notin \partial G(x^0)$.

- D'une part remarquons que $\partial L(x^0) = Q$. En effet puisque $\forall\ i = 1, \cdots, m$ les fonctions f_i sont convexes, $y \geq 0$ et $\cap_{i=1}^m ri(\text{dom } f_i) \neq \emptyset$, nous déduisons d'une propriété du sous-différentiel (proposition A.IV.15) que

$$\partial L(x^0) = \sum_{i=1}^{m} y_i \partial f_i(x^0) + N_{(\cap_{i=1}^m \text{dom } f_i)}(x^0)\ .$$

  Par ailleurs

$$N_{(\text{dom } G)}(x^0) = N_{(\cap_{i=1}^m \text{dom } f_i)}(x^0)$$

  car f_0 est différentiable sur $I\!R^m$, d'où dom $f_0 = I\!R^m$.
  Par conséquent

$$\partial L(x^0) = \sum_{i=1}^{m} y_i \partial f_i(x^0) + N_{(\text{dom } G)}(x^0) \tag{4.63}$$

  La thèse suit alors de la définition de Q.

- D'autre part, montrons que $\forall\ w \in I\!R^n \setminus Q, w \notin \partial G(x^0)$.
  Soit un vecteur $w \in I\!R^n \setminus Q$ arbitraire.
  Puisque $Q = \partial L(x^0), w \notin \partial L(x^0)$, et donc par la proposition A.IV.10,

$$\exists\ d' \in I\!R^n \text{ tel que } < w, d' > > L'(x^0; d')\ .$$

  Notons que s.p.d.g. $||d'|| = 1$.
  Or par le pas 2 $L'(x^0; d') = G'(x^0; d')$, d'où

$$\exists\ d' \in I\!R^n\ ||d'|| = 1 \text{ tel que } < w, d' > > G'(x^0; d')\ .$$

De plus comme la dérivée directionnelle de G en x^0 est la fonction d'appui du sous-différentiel de G en ce point, c'est-à-dire

$$G'(x^0; d') = \sup\{< x^*, d' > \text{ tel que } x^* \in \partial G(x^0)\} ,$$

nous concluons que $w \notin \partial G(x^0)$.
Ce qu'il fallait démontrer. ∎

Prouvons maintenant le théorème de convergence associé aux problèmes (PI) et (DI).

Preuve du théorème 4.5 :

Pas 1 : Prouvons d'abord la convergence de la suite $\{p^k\}$.

Par le théorème 4.4, nous savons que $p^{k+1} = ((\nabla h + \lambda_k \partial(-d) \circ \nabla h)(p^k)$.
Or il suit de la proposition A.IV.9 que

$$\overline{D(\partial(-d))} \subseteq \overline{\text{dom } (-d)} \subseteq \overline{\Omega}^+ ,$$

ainsi, comme par hypothèse $\overline{\Omega}^+ \subseteq S$, nous obtenons que $\overline{D(\partial(-d))} \subseteq S$.
Par conséquent il résulte du théorème 2.7, appliqué à l'opérateur maximal monotone $\partial(-d)$, que la suite $\{p^k\}$ converge vers un zéro de $\partial(-d)$, c'est-à-dire une solution du problème dual (DI), s'il en existe une. Dans le cas contraire, la suite $\{p^k\}$ est non bornée.

Pas 2 : Supposons à présent que le problème dual (DI) admet au moins une solution.

Soient $\overline{p}$ la limite de la suite $\{p^k\}$ et $\overline{x}$ une valeur d'adhérence de la suite $\{x^k\}$.
S.p.d.g., nous notons $\{x^k\}$ la sous-suite qui converge vers $\overline{x}$.
Nous désirons prouver que $\overline{x}$ est une solution du problème (PI). Dans ce but, nous montrerons que $\overline{x}$ vérifie les *conditions de Kuhn Tucker* associées au problème (PI).

Considérons l'opérateur P_k défini sur $I\!R^m$ par $P_k(u) = \dfrac{1}{\lambda_k} h^{*+}(\nabla h(p^k) + \lambda_k u)$. (4.64)

1. Dans un premier temps, montrons que $\overline{x}$ est admissible pour (PI).
   - D'une part, puisque la suite $\{x^k\}$ est contenue dans l'ensemble fermé X, il est évident que $\overline{x} \in X$.
   - D'autre part $g_i(\overline{x}) \leq 0 \quad i = 1, \cdots, m$.
     En effet, en vertu du lemme 4.6, nous savons que

$$g(x^k) - \frac{1}{\lambda_k}[\nabla h(p^{k+1}) - \nabla h(p^k)] \in N_{\overline{\Omega}^+}(p^{k+1}) . \tag{4.65}$$

Or la fonction h est continûment différentiable sur S et inf $_{k\geq 0}$ $\{\lambda_k\} > 0$, dès lors

$$\lim_{k\to\infty} \frac{1}{\lambda_k}(\nabla h(p^{k+1}) - \nabla h(p^k)) = 0 \ . \tag{4.66}$$

De plus, puisque par hypothèse les fonctions g_i $i = 1, \cdots, m$ sont continues sur X, il suit que

$$\lim_{k\to\infty} g(x^k) = g(\overline{x}) \ . \tag{4.67}$$

Ainsi en vertu des relations (4.66) et (4.67), nous obtenons que

$$\lim_{k\to\infty} \{g(x^k) - \frac{1}{\lambda_k}[\nabla h(p^{k+1}) - \nabla h(p^k)]\} = g(\overline{x}) \ . \tag{4.68}$$

Par ailleurs $N_{\overline{\Omega}^+}$ est le sous-différentiel de la fonction propre, convexe et s.c.i. δ^+ (cfr proposition A.V.2). Il s'agit donc d'un opérateur maximal monotone (cfr exemple 2.5). Par conséquent, nous déduisons de (4.65), (4.68) et du caractère fermé des opérateurs maximaux monotones (cfr proposition 2.4) que

$$g(\overline{x}) \in N_{\overline{\Omega}^+}(\overline{p}) \ , \tag{4.69}$$

ce qui signifie $g_i(\overline{x}) \leq 0$ $i = 1, \cdots, m$ car $N_{\overline{\Omega}^+} = \overline{\Omega}^-$.

2. En second lieu, montrons que

$$\overline{p} \geq 0 \quad \text{et} \quad < \overline{p}, g(\overline{x}) > \ = \ 0 \tag{4.70}$$

- Puisque $p^{k+1} = \nabla h^{*+}(\nabla h(p^k) + \lambda_k g(x^k))$, il suit immédiatement de la définition de l'opérateur P_k (cfr relation (4.64)) que $p^{k+1} = \nabla P_k(g(x^k))$.

  Or par le lemme 4.4 l'opérateur P_k est croissant, dès lors $p^{k+1} \geq 0 \ \forall \ k \geq 0$. (4.71)

  Ainsi, comme la suite $\{p^k\}$ est contenue dans l'ensemble fermé $\overline{\Omega}^+$, il suit que $\overline{p} \in \overline{\Omega}^+$.

- Définissons maintenant l'ensemble $J := \{j = 1, \cdots, m$ tel que $g_j(\overline{x}) < 0\}$.
  Puisque par (4.69) $g(\overline{x}) \in N_{\overline{\Omega}^+}(\overline{p})$, nous obtenons que $< g(\overline{x}), z - \overline{p} > \leq 0$ pour tout $z \in \overline{\Omega}^+$.
  Autrement dit,

  $$\sum_{j\in J} g_j(\overline{x})(z_j - \overline{p}_j) \leq 0 \quad \text{pour tout } z_j \geq 0 \ j \in J \ .$$

  Ainsi, comme $\overline{p} \in \overline{\Omega}^+$, nous concluons que $\overline{p}_j = 0$ pour tout $j \in S$.
  D'où $< \overline{p}, g(\overline{x}) >= 0$.

3. Pour terminer prouvons que $\overline{x}$ $\in$argmin $_{x\in I\!R^n}$ $\{L(x,\overline{p})\}$ où la fonction lagrangienne associée à (PI) est définie par $L(x,\overline{p}) = f_X(x) + \sum_{i=1}^m \overline{p}_i g_i(x)$.

Dans ce but, nous vérifierons d'abord que

$$0 \in \partial f(x^k) + N_X(x^k) + \sum_{i=1}^{m} p_i^{k+1} \partial g_i(x^k) \quad \forall\, k \geq 0$$

Ensuite nous montrerons que pour tout $p \in \overline{\Omega}^+$, l'opérateur $\partial f + N_X + \sum_{i=1}^m p_i \partial g_i$ est maximal monotone.
Enfin, nous basant sur le caractère fermé des opérateurs maximaux monotones (cfr proposition 4.4), nous concluerons que $0 \in \partial f(\overline{x}) + N_X(\overline{x}) + \sum_{i=1}^m \overline{p}_i g_i(\overline{x})$, autrement dit, $\overline{x}$ $\in$argmin $_{x\in I\!R^n}$ $\{f_X(x) + \sum_{i=1}^m \overline{p}_i g_i(x)\}$.

a) Vérifions d'abord que

$$0 \in \partial f(x^k) + N_X(x^k) + \sum_{i=1}^{m} p_i^{k+1} \partial g_i(x^k) \quad \forall\, k \geq 0\,. \tag{4.72}$$

- Comme

$$x^k \in \text{argmin}_{\,x\in I\!R^n} \{f_X(x) + \frac{1}{\lambda_k} h^{*+}(\nabla h(p^k) + \lambda_k g(x))\}$$

il suit que

$$0 \in \partial_x[f(.) + \delta_X(.) + P_k(g(.))](x^k)\,. \tag{4.73}$$

Or $ri(\text{dom } f) \cap ri(\text{dom } \delta_X) \cap ri(\text{dom } (P_k \circ g)) \neq \emptyset$. En effet, par hypothèse on a $X \subseteq$dom f et $X \subseteq$dom g.
Dès lors, il suit des propositions A.II.4 et A.II.8 que

$$ri(X) \neq \emptyset \text{ et } ri(X) \subseteq ri(\text{dom } f) \cap ri(\text{dom } \delta_X) \cap ri(\text{dom } g)\,,$$

$$\text{d'où } ri(\text{dom } f) \cap ri(\text{dom } \delta_X) \cap ri(\text{dom } g) \neq \emptyset\,. \tag{4.74}$$

Or dom $(P_k \circ g) =$dom g car dom $P_k = I\!R^m$ (cfr lemme 4.1), ainsi,

$$ri(\text{dom } f) \cap ri(\text{dom } \delta_X) \cap ri(\text{dom } (P_k \circ g)) \neq \emptyset\,. \tag{4.75}$$

Par conséquent, nous déduisons de la propriété de sommation du sous-différentiel (cfr proposition A.IV.12) que

$$\partial f(x^k) + N_X(x^k) + \partial(P_k \circ g)(x^k)\,. \tag{4.76}$$

- Evaluons maintenant $\partial(P_k \circ g)(x^k)$.
D'une part l'opérateur P_k est différentiable et croissant sur $I\!R^m$ (cfr lemme 4.4).

D'autre part les fonctions g_i $i = 1, \cdots, m$ sont propres, convexes et $\cap_{i=1}^m ri(\text{dom}\ \ g_i) \neq \emptyset$ (cfr relation 4.73).
Dès lors il suit de la proposition 4.16 que

$$\begin{aligned}
&\partial(P_k(g(x^k))) = \sum_{i=1}^m [\nabla P_k(g(x^k))]_i \partial g_i(x^k) + N_{(\cap_{i=1}^m \text{dom}\ g_i)}(x^k) \\
\Leftrightarrow\quad &\partial(P_k(g(x^k))) = \sum_{i=1}^m [\nabla h^{*+}(\nabla h(p^k) + \lambda_k g(x^k))]_i \partial g_i(x^k) + N_{(\cap_{i=1}^m}\ \text{dom}\ \ g_i) \\
\Leftrightarrow\quad &\partial(P_k(g(x^k))) = \sum_{i=1}^m p_i^{k+1} \partial g_i(x^k) + N_{(\cap_{i=1}^m \text{dom}\ g_i)}(x^k)\ . \qquad (4.77)
\end{aligned}$$

- Remplaçant $\partial(P_k \circ g)(x^k)$ par sa valeur (4.77) dans (4.76) nous obtenons que

$$0 \in \partial f(x^k) + N_X(x^k) + \sum_{i=1}^m p_i^{k+1} \partial g_i(x^k) + N_{(\cap_{i=1}^m \text{dom}\ g_i)}(x^k)\ .$$

La thèse suit alors d'une propriété du cône normal. En effet, par hypothèse $X \subseteq \cap_{i=1}^m(\text{dom}\ \ g_i)$ d'où

$$N_X + N_{(\cap_{i=1}^m \text{dom}\ g_i)} = N_X\ . \qquad (4.78)$$

(cfr proposition A.V.3).

b) Prouvons à présent que pour tout $p \in \overline{\Omega}^+$, l'opérateur

$$\partial f + N_X + \sum_{i=1}^m p_i \partial g_i \text{ est maximal monotone.} \qquad (4.79)$$

En effet, nous savons déjà par la relation (4.74) que

$$ri(\text{dom}\ \ f) \cap ri(\text{dom}\ \ \delta_X) \cap ri(\cap_{i=1}^m(\text{dom}\ \ g_i)) \neq \emptyset$$

Ainsi, la propriété de sommation du sous-différentiel (cfr proposition A.IV.12) nous assure que pour tout $x \in (\text{dom}\ \ f) \cap \text{dom}\ \ (\delta_X) \cap (\text{dom}\ \ g)$,

$$\partial(f + \delta_X + \sum_{i=1}^m p_i g_i)(x) = \partial f(x) + N_X(x) + \partial(\sum_{i=1}^m p_i g_i)(x)\ .$$

De plus puisque par la relation (4.74) $\cap_{i=1}^m ri(\text{dom}\ \ g_i) \neq \emptyset$, il suit d'une propriété du sous-différentiel (cfr proposition A.IV.15) que

$$\partial(\sum_{i=1}^m p_i g_i)(x) = \sum_{i=1}^m p_i \partial g_i(x) + N_{(\cap_{i=1}^m \text{dom}\ g_i)}(x)\ ,$$

dès lors

$$\partial(f+\delta_X+\sum_{i=1}^{m} p_i g_i)(x) = \partial f(x)+N_X(x)+\sum_{i=1}^{m} p_i\partial g_i(x)+N_{(\cap_{i=1}^m \text{dom } g_i)}(x) \ . \quad (4.80)$$

Or nous savons par (4.78) que $N_X + N_{(\cap_{i=1}^m \text{dom } g_i)} = N_X$,
par conséquent, nous obtenons que

$$\partial(f + \delta_X + \sum_{i=1}^{m} p_i g_i)(x) = \partial f(x) + N_X(x) + \sum_{i=1}^{m} p_i\partial g_i(x) \ .$$

D'autre part, comme les fonctions f, δ_X et $p_i g_i (p_i \geq 0)$ sont propres, convexes et s.c.i., il résulte de l'exemple 2.5 que l'opérateur $\partial(f + \delta_X + \sum_{i=1}^m p_i g_i)$ est maximal monotone.
La thèse suit alors de l'égalité des deux opérateurs

$$\partial(f + \delta_X + \sum_{i=1}^{m} p_i g_i) \text{ et } \partial f + N_X + \sum p_i\partial g_i \ .$$

c) Finalement, montrons que $\overline{x} \in$argmin $_{x\in I\!R^n}$ $\{f_X(x) + \sum_{i=1}^m \overline{p}_i g_i(x)\}$.

Remarquons dès à présent que les opérateurs $\partial f + N_X + \sum_{i=1}^m p_i^{k+1}\partial g_i$ et $\partial f + N_X + \sum_{i=1}^m \overline{p}_i\partial g_i$ sont maximaux monotones car par les relations (4.70) et (4.71) $p^k \geq 0 \quad \forall\, k \geq 0$ et $\overline{p} \geq 0$.

Ainsi, comme

$$0 \in \partial f(x^k) + N_X(x^k) + \sum_{i=1}^{m} p_i^{k+1}\partial g_i(x^k) \ ,$$

nous déduisons de la monotonie de l'opérateur $\partial f + N_X + \sum_{i=1}^m p_i^{k+1}\partial g_i(x^k)$ que $\forall\, x, y, z, \gamma^1, \cdots, \gamma^m \in I\!R^n$ tels que $y \in \partial f(x), z \in N_X(x)$ et
$\gamma^i \in \partial g_i(x) \quad i = 1, \cdots, m$
on a

$$< x - x^k, y + z + \sum_{i=1}^{m} p_i^{k+1}\gamma^i - 0 > \ \geq 0 \ .$$

Fixons maintenant de tels $\hat{x}, \hat{y}, \hat{z}, \hat{\gamma}^1 \cdots \hat{\gamma}^m$. Lorsque k tend vers l'infini nous obtenons que

$$< \hat{x} - \overline{x}, \hat{y} + \hat{z} + \sum_{i=1}^{m} \overline{p}_i\hat{\gamma}^i - 0 > \geq 0 \ .$$

Par conséquent, comme les vecteurs $\hat{x}, \hat{y}, \hat{z}, \hat{\gamma}^1 \cdots \hat{\gamma}^m$ sont arbitraires, il résulte du caractère maximal de l'opérateur $\partial f + N_X + \sum_{i=1}^m \overline{p}_i \partial g_i$ (cfr définition 2.9) que

$$0 \in \partial f(\overline{x}) + N_X(\overline{x}) + \sum_{i=1}^m \overline{p}_i \partial g_i(\overline{x}) ,$$

autrement dit

$$\overline{x} \in \text{argmin}_{x \in I\!R^n} \{f_X(x) + \sum_{i=1}^m \overline{p}_i g_i(x)\} .$$

4. En conclusion, il résulte de 1), 2), 3) que $\overline{x}$ vérifie les *conditions de Kuhn Tucker* associées au problème de minimisation convexe (PI). En effet,

   $\overline{x}$ est admissible,

   $< g(\overline{x}), \overline{p} >= 0$

   et $\overline{x} \in \text{argmin}_{x \in I\!R^n} \{L(x, \overline{p})\}$ où $\overline{p}$ est une solution du dual (DI).

   Par conséquent $\overline{x}$ est solution du problème (PI). ∎

Dans la suite de ce paragraphe, consacré aux problèmes avec contraintes d'inégalités, nous écrirons l'itération de la méthode des multiplicateurs utilisant des fonctions de Bregman h particulières.

4.4.5 Cas particulier : la fonction de Bregman h est séparable

Soit h une *fonction de Bregman séparable*, c'est-à-dire

$$h(u) = \sum_{i=1}^m h_i(u_i)$$

où $h_1, \cdots, h_m$ sont des fonctions de Bregman de zone respective $S_i \supset [0, \infty)$ $i = 1, \cdots, m$.
Il est donc immédiat que l'hypothèse $\overline{\Omega}^+ \subseteq S$ est vérifiée.

Remarquons dès à présent que

$$\nabla h_i(0) \text{ est un minimum global de } h_i^* \text{ sur } I\!R \quad \forall\, i = 1, \cdots, m .$$

En effet, puisque les fonctions h_i^* sont propres, convexes et différentiables sur dom $h_i^* = S_i$ (lemme 4.1), il résulte de la proposition 4.2 que

$$\nabla h_i^*(\nabla h_i(0)) = 0 \Leftrightarrow \nabla h_i(0) = \nabla h_i(0) .$$

Ce qui est évidemment vrai.
La thèse découle alors du théorème A.VII.3.

Evaluons maintenant la fonction conjuguée monotone de h, ainsi que son gradient. Notons que par convention pour tout $x, y \in I\!R^m$, $\underline{\max}\{x, y\}$ désigne le vecteur dont chaque composante est le maximum entre x_i et y_i $\; i = 1, \cdots, m$.

Proposition 4.7 :
Soit h une fonction de Bregman séparable. Alors $h^{*+}(u) = h^*(\underline{\max}\{u, \nabla h(0)\})$.

Preuve de la proposition 4.17 :

1. La définition de la conjuguée monotone nous assure que :

$$\begin{aligned} & h^{*+}(u) = \sup_{p \geq 0}\{\sum_{i=1}^{m} p_i u_i - \sum_{i=1}^{m} h_i(p_i)\} \\ \Leftrightarrow \quad & h^{*+}(u) = \sum_{i=1}^{m} \sup_{p_i \geq 0}\{p_i u_i - h_i(p_i)\} \\ \Leftrightarrow \quad & h^{*+}(u) = \sum_{i=1}^{m} h_i^{*+}(u_i) \end{aligned} \tag{4.81}$$

2. Evaluons maintenant $h_i^{*+}(u_i)$ pour $i \in \{1, \cdots, m\}$ arbitraire.
En vertu de la proposition 4.10, nous savons que

$$h_i^{*+}(u_i) = \inf{}_{w_i \geq u_i} h_i^*(w_i) \; .$$

De deux choses l'une :

(a) Si $u_i \leq \nabla h_i(0)$ alors

$$h_i^{*+}(u_i) = \inf{}_{w_i \geq u_i} h_i^*(w_i) = h_i^*(\nabla h_i(0)) \tag{4.82}$$

car $\nabla h_i(0)$ est un minimum global de h_i^* sur $I\!R$.

(b) Si $u_i > \nabla h_i(0)$ alors

$$h_i^{*+}(u_i) = \inf{}_{w_i \geq u_i} h_i^*(w_i) = h_i^*(u_i) \tag{4.83}$$

En effet, par une propriété des fonctions convexes (cfr proposition A.I.13) on a que pour tout $w_i \geq u_i > \nabla h_i(0)$

$$\frac{h_i^*(w_i) - h_i^*(\nabla h_i(0))}{w_i - \nabla h_i(0)} \geq \frac{h_i^*(u_i) - h_i^*(\nabla h_i(0))}{u_i - \nabla h_i(0)}$$

$$\Leftrightarrow [h_i^*(w_i) - h_i^*(\nabla h_i(0))][u_i - \nabla h_i(0)] \geq [h_i^*(u_i) - h_i^*(\nabla h_i(0))][w_i - \nabla h_i(0)] \tag{4.84}$$

D'autre part $\forall\ w_i \geq u_i > \nabla h_i(0)\ \ (w_i - \nabla h_i(0)) \geq (u_i - \nabla h_i(0)) \geq 0$

$$\text{d'où } [h_i^*(u_i) - h_i^*(\nabla h_i(0))][w_i - \nabla h_i(0)] \geq [h_i^*(u_i) - h_i^*(\nabla h_i(0))][u_i - \nabla h_i(0)] \tag{4.85}$$

Ainsi, il résulte des relations (4.84), (4.85), et de la positivité stricte de $(u_i - \nabla h_i(0))$ que

$$\begin{aligned} & [h_i^*(w_i) - h_i^*(\nabla h_i(0))][u_i - \nabla h_i(0)] \geq [h_i^*(u_i) - h_i^*(\nabla h_i(0))][u_i - \nabla h_i(0)] \\ \Leftrightarrow\ & h_i^*(w_i) - h_i^*(\nabla h_i(0)) \geq h_i^*(u_i) - h_i^*(\nabla h_i(0)) \\ \Leftrightarrow\ & h_i^*(w_i) \geq h_i^*(u_i) \end{aligned}$$

Par conséquent, nous obtenons que

$$\forall\ w_i \geq u_i > \nabla h_i(0),\ \ h_i^*(w_i) \geq h_i^*(u_i) \tag{4.86}$$

3. En conclusion, nous désuisons des relations (4.81), (4.82), (4.83) que

$$\begin{aligned} & h^{*+}(u) = \sum_{i=1}^{m} h_i^{*+}(u_i) \\ \Leftrightarrow\ & h^{*+}(u) = \sum_{i=1}^{m} h_i^* \max(u_i, \nabla h_i(0)) \\ \Leftrightarrow\ & h^{*+}(u) = h^*(\underline{\max}\{u, \nabla h(0)\}) \end{aligned}$$

■

Par un raisonnement analogue, nous pouvons prouver la proposition suivante :

Proposition 4.18 :

Soit h une fonction de Bregman séparable. Alors

$$\nabla h_i^{*+}(u_i) = \max\{0, \nabla h_i(0)\} \quad i = 1, \cdots, m\ .$$

Nous disposons maintenant de tous les résultats nécessaires pour écrire l'algorithme de *la méthode des multiplicateurs utilisant une fonction de Bregman séparable.*

<u>**Algorithme 4.7 :**</u>

Soient $\{\lambda_k\}_{k=0}^{\infty}$ une suite de scalaires strictement positifs; h une fonction de Bregman séparable telle que $\Omega^+ \subseteq R(\nabla h)$. A l'itération k, soient $p^k \in \overline{\Omega}^+$ et $\lambda_k > 0$ connus.

1. Résoudre le problème de minimisation

$$x^k \in \text{argmin}_{\,x\in \mathbb{R}^n} \{f_X(x) + \frac{1}{\lambda_k} h^*(\underline{\max}\{\nabla h(0), \nabla h(p^k) + \lambda_k g(x)\})\} \,.$$

2. calculer l'itération des multiplicateurs

$$p^{k+1} = \underline{\max}\{0, \nabla h^*(\nabla h(p^k) + \lambda_k g(x^k))\} \,.$$

4.4.6 Cas particulier $h(u) = \frac{1}{2}||u||^2$

Il est évident que la fonction h définie de $\mathbb{R}^m$ dans $\mathbb{R}$ par $h(u) = \frac{1}{2}||u||^2$ est une fonction de Bregman séparable de zone $\mathbb{R}^m$. De plus $R(\nabla h) = \mathbb{R}^m$ puisque $\nabla h(u) = u \quad \forall\, u \in \mathbb{R}^m$.

Soit $p^0 \in \overline{\Omega}^+$. Comme pour tout $u \in \mathbb{R}^m$ $h^*(u) = \frac{1}{2}||u||^2$ (cfr exemple A.VI.3), la méthode des multiplicateurs génère une suite $\{(x^k, p^k)\}$ telle que

$$\left\{\begin{array}{l} x^k \in \text{argmin}_{\,x\in \mathbb{R}^n} \{f_X(x) + \frac{1}{\lambda_k} h^*(\underline{\max}\{\nabla h(0), \nabla h(p^k) + \lambda_k g(x)\})\} \\ p^{k+1} = \underline{\max}\{0, \nabla h^*(\nabla h(p^k) + \lambda_k g(x^k))\} \end{array}\right.$$

$$\Leftrightarrow \left\{\begin{array}{l} x^k \in \text{argmin}_{\,x\in \mathbb{R}^n} \{f_X(x) + \frac{1}{2\lambda_k}||\underline{\max}(0, p^k + \lambda_k g(x))||^2\} \\ p^{k+1} = \underline{\max}\{0, p^k + \lambda_k g(x^k)\} \end{array}\right.$$

$$\Leftrightarrow \left\{\begin{array}{l} x^k \in \text{argmin}_{\,x\in \mathbb{R}^n} \{f_X(x) + \frac{1}{2\lambda_k} \sum_{i=1}^m \{\max(0, p_i^k + \lambda_k g_i(x))\}^2\} \\ p_i^{k+1} = \max\{0, p_i^k + \lambda_k g_i(x^k)\} \quad i = 1, \cdots, m \end{array}\right.$$

Notre méthode est donc équivalente à celle du *lagrangien augmenté classique* pour le problème (PI) (algorithme 4.2).

4.4.7 Cas particulier $h(u) = \sum_{i=1}^{m} (u_i \ln u_i - u_i)$

Ecrivons l'itération de la méthode des multiplicateurs utilisant la fonction de Bregman h de zone $S = \Omega^+$ définie par $h(u) = \sum_{i=1}^{m} (u_i \ln u_i - u_i)$
où $\Omega^+ = \{u \in I\!R^m$ tel que $u > 0\}$ (cfr exemple 2.9).

Notons que la fonction h ainsi définie n'étant pas séparable, nous ne pouvons pas appliquer l'algorithme 2.7.

Evaluons donc la fonction conjuguée monotone de h.
En vertu de la proposition 4.10, nous savons que

$$\forall\ u \in \text{dom}\ \ h^{*+}, h^{*+}(u) = \inf_{w \geq u} h^*(w)\ .$$

Or par [19, p. 142] $h^*(u) = \sum_{i=1}^{m}\ h_i^*(u_i)$ où la fonction h_i est définie par
$h_i(u_i) = u_i \ln u_i - u_i$.
D'où en vertu de l'exemple A.VI.3 (2) on a $h^*(u) = \sum_{i=1}^{m}\ e^{u_i}$.
Par conséquent, $\inf_{w \geq u}\ h^*(w) = h^*(u)$ c'est-à-dire

$$\forall\ u \in \text{dom}\ \ h^{*+} \quad h^{*+}(u) = h^*(u) = \sum_{i=1}^{m}\ e^{u_i}$$

Par ailleurs $\nabla h(u) = (\ln u_i)_{i=1}^{m}$ et $[\nabla h^{*+}(u)]_i = e^{u_i} \quad i = 1, \cdots, m$.

Ainsi, nous obtenons les équivalences suivantes :

$$\begin{cases} x^k \in \text{argmin}_{\ x \in I\!R^n}\ \{f x(x) + \frac{1}{\lambda_k} h^{*+}(\nabla h(p^k) + \lambda_k g(x))\} \\ p^{k+1} = \nabla h^{*+}(\nabla h(p^k) + \lambda_k g(x^k)) \end{cases}$$

$$\Leftrightarrow \begin{cases} x^k \in \text{argmin}_{\ x \in I\!R^n}\ \{f x(x) + \frac{1}{\lambda_k} \sum_{i=1}^{m}\ e^{\ln p_i^k + \lambda_k g_i(x)}\} \\ p_i^{k+1} = e^{\ln p_i^k + \lambda_k g_i(x^k)} \qquad i = 1, \cdots, m \end{cases}$$

$$\Leftrightarrow \begin{cases} x^k \in \text{argmin}_{\ x \in I\!R^n}\ \{f x(x) + \frac{1}{\lambda_k} \sum_{i=1}^{m}\ p_i^k e^{\lambda_k g_i(x)}\} \\ p_i^{k+1} = p_i^k e^{\lambda_k g_i(x^k)} \quad i = 1, \cdots, m \end{cases}$$

Ces deux formules sont bien connues sous le nom de *méthode exponentielle des multiplicateurs* (algorithme 4.3).

Malheureusement, comme $S = \Omega^+$, nous ne pouvons pas appliquer notre théorie à cette méthode. En effet nous ne disposons pas de l'hypothèse $\overline{\Omega}^+ \subseteq S$. Plus précisément, rappelons que la suite des multiplicateurs $\{p^k\}$ est obtenue en appliquant l'algorithme du point proximal non linéaire (algorithme 2.1) à l'opérateur maximal monotone $\partial(-d)$ (où d est la fonction duale associée au problème (PI)).
Ainsi, comme dans notre cas, l'hypothèse $\overline{D(\partial fd))} \subseteq S$ n'est pas nécessairement vérifiée, nous ne pouvons pas utiliser le théorème 2.7 pour prouver la convergence de la suite $\{p^k\}$ vers une solution du problème dual (DI) (cfr théorème 4.5 pas 1).
D'où l'intérêt d'affaiblir l'hypothèse $\overline{D(T)} \subseteq S$ dans l'analyse de convergence de l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman (nous nous référons ici à la remarque 2.2 du paragraphe 2.4.4).

Notons que dans leur article [8], Gong Chen et Marc Teboulle ont défini la suite des multiplicateurs $\{p^k\}$ en appliquant l'algorithme de minimisation proximale avec fonctions de Bregman (cfr algorithme 3.1) au problème dual (DI). Cette approche est bien sûr équivalente à la nôtre. En effet, le théorème 3.1 nous assure que

$$p^{k+1} = ((\nabla h + \lambda_k \partial(-d))^{-1} \circ \nabla h)(p^k) \Leftrightarrow p^{k+1} = \text{argmin}_{\,p \geq 0} \{-d(p) + \frac{1}{\lambda_k} D_h(p, p^k)\} \,.$$

Ainsi, ayant prouvé la convergence de l'algorithme de minimisation proximale sous l'hypothèse plus faible $ri(\text{dom}\ \ d) \subseteq S$ (cfr remarque 3.1), ils ont pu prouver la convergence de la méthode des multiplicateurs.

Citons leur théorème :

Théorème :
Soient h la fonction de Bregman de zone Ω^+ définie par $h(p) = \sum_{i=1}^m (p_i \ln p_i - p_i)$ et $\{\lambda_k\}$ une suite de scalaires strictement positifs tels que $\inf_{\,k \geq 0} \ \{\lambda_k\} > 0$.
Considérons le problème de minimisation convexe avec contraintes d'inégalités (PI), et supposons que $ri(\text{dom}\ \ f) \cap (\cap_{i=1}^m ri(\text{dom}\ \ g_i)) \neq \emptyset$.
Alors la suite des multiplicateurs de Lagrange $\{p^k\}$ définie par

$$p^{k+1} = \text{argmin}_{\,p \geq 0} \ \{-d(p) + \frac{1}{\lambda_k} D_h(p, p^k)\}$$

converge vers une solution du dual (DI) s'il en existe une. De plus dans ce cas, si les fonctions $g_i \quad i = 1, \cdots, m$ sont continues, alors toute valeur d'adhérence de la suite $\{x^k\}$ est une solution du problème (PI).

Preuve : cfr [8], théorème 4.2.

4.4.8 Cas particulier : $h(u) = \sum_{i=1}^m (u_i + 1)[\ln(u_i + 1) - 1)]$

Pour terminer réécrivons l'itération de la méthode des multiplicateurs utilisant la fonction de Bregman h de zone $S := \{p \in I\!R^m$ tel que $p_i > -1 \quad i = 1, \cdots, m\}$ définie par $h(u) = \sum_{i=1}^m (u_i + 1)[\ln(u_i + 1) - 1]$.

Comme $\overline{\Omega}^+ \subseteq S$ nous pouvons appliquer notre théorie à cette méthode.

Il est évident que $\nabla h(p) = \ln(p_i + 1)_{i=1}^m$.

Par ailleurs $h^*(u) = \sum_{i=1}^m (e^{u_i} - u_i)$. De plus, nous prenons comme convention que le vecteur $\mathbf{e}^u = (e^{u_1} \cdots e^{u_m})$ et que le vecteur $\mathbf{1} = (1 \ldots 1)$.

Considérons un vecteur $p^0 \in \overline{\Omega}^+$. Dès lors la méthode des multiplicateurs génère une suite $\{(x^k, p^k)\}$ telle que

$$\begin{cases} x^k \in \text{argmin}_{x \in I\!R^n} \{f_X(x) + \frac{1}{\lambda_k} \sum_{i=1}^m (\max\{1, e^{\ln(p_i^k+1)+\lambda_k g_i(x)} - \ln(p_i^k + 1) - \lambda_k g_i(x)\}\})\} \\ p^{k+1} = \underline{\max}\{0, \mathbf{e}^{(\nabla h(p^k)+\lambda_k g(x^k))} - \mathbf{1}\} \end{cases}$$

$$\Leftrightarrow \begin{cases} x^k \in \text{argmin}_{x \in I\!R^n} \{f_X(x) + \frac{1}{\lambda_k} \sum_{i=1}^m (\max\{1, (p_i^k + 1)e^{\lambda_k g_i(x)} - \ln(p_i^k + 1) - \lambda_k g_i(x)\}\})\} \\ p^{k+1} = \underline{\max}\{0, (p^k + \mathbf{1})\mathbf{e}^{\lambda_k g(x^k))} - \mathbf{1}\} \end{cases}$$

L'algorithme résultant possède un désavantage non négligeable par rapport à la méthode exponentielle classique des multiplicateurs classique (cfr paragraphe 4.4.7). En effet, les dérivées secondes de la fonction $f(x) = \frac{1}{\lambda_k} \sum_{i=1}^m (\max\{1, (p_i^k + 1)e^{\lambda_k g_i(x)} - \ln(p_i^k + 1) - \lambda_k g_i(x)\}\})$ étant discontinues, nous ne pouvons pas utiliser la méthode de Newton pour rechercher x^k.

4.5 Problème de minimisation convexe avec contraintes d'égalités et d'inégalités

Dans ce chapitre, nous avons étudié une nouvelle méthode des multiplicateurs. Pour des raisons de clarté, nous avons séparé les problèmes avec contraintes d'égalités et d'inégalités. Combinons maintenant les deux approches.

Considérons le problème

$$(P) \qquad \left\{ \begin{array}{ll} \text{Minimiser} & f(x) \\ \text{s.c.} & g_i(x) \leq 0 \quad i = 1, \cdots, m \\ & Ax = b \\ & x \in X \end{array} \right.$$

où $f; g_i \quad i = 1, \cdots, m; A; b$ et X vérifient les hypothèses des problèmes (PI) et (PE).

Soient h_1 et h_2 deux fonctions de Bregman définies, respectivement de $\overline{S_1} \subseteq I\!R^m$ et $\overline{S_2} \subseteq I\!R^m$ dans $I\!R$.
Supposons également que

$$\begin{array}{lll} S_1 = I\!R^m & \text{et} & \nabla h_1 = I\!R^m \\ S_2 \supseteq \overline{\Omega}^+ & \text{et} & \Omega^+ \subseteq R(\nabla h_2) \, . \end{array}$$

Etant donné $y^0 \in I\!R^m$ et $p^0 \in \overline{\Omega}^+$, l'algorithme de la *méthode des multiplicateurs utilisant des fonctions de Bregman adaptée au problème* (P) génère une suite $\{(x^k, y^k, p^k)\} \subseteq I\!R^n \times I\!R^m \times \overline{\Omega}^+$ telle que

$$\left\{ \begin{array}{l} x^k \in \text{argmin}_{\, x \in I\!R^n} \{ f_X(x) + \frac{1}{\lambda_k} h_1^*(\nabla h_1(y^k) + \lambda_k(Ax - b)) \\ \qquad\qquad\qquad + \frac{1}{\lambda_k} h_2^{*+}(\nabla h_2(p^k) + \lambda_k g(x)) \} \\ y^{k+1} = \nabla h_1^*(\nabla h_1(y^k) + \lambda_k(Ax^k - b)) \\ p^{k+1} = \nabla h_2^*(\nabla h(p^k) + \lambda_k g(x^k)) \end{array} \right.$$

Les propriétés de convergence de cette suite sont données par les théorèmes 4.3 et 4.5.

Chapitre 5

Méthode proximale des multiplicateurs utilisant des fonctions de Bregman

Dans ce chapitre, nous appliquons l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman (algorithme 2.1) à la formulation "primal-dual" d'un problème d'optimisation convexe. Nous obtenons ainsi *la première méthode proximale non quadratique des multiplicateurs.*

Afin de faciliter la comparaison avec le cas classique [17, paragraphe 5], nous nous limiterons aux problèmes avec contraintes d'inégalités, c'est-à-dire :

$$(\mathrm{PI}) \qquad \begin{cases} \text{Minimiser} & f(x) \\ \quad\text{s.c.} & g_i(x) \leq 0 \quad i = 1, \cdots, m \\ & x \in X \end{cases}$$

où $f, g_i \quad i = 1, \cdots,$ sont des fonctions propres, convexes et s.c.i définies de $I\!R^n$ dans $I\!R \cup \{+\infty\}$ et X un ensemble convexe, fermé, non vide de $I\!R^n$.
Nous supposons également que :

- les fonctions f et $g_i \quad i = 1, \cdots, m$ ont une valeur finie sur X
- l'ensemble dom $f \cap \{x \in X$ tel que $g_i(x) \leq 0 \quad i = 1, \cdots, m\} \neq \emptyset$
- l'hypothèse de qualification de contraintes - $\exists \quad \hat{x} \in X$ tel que $g_i(\hat{x}) < 0 \quad i = 1, \cdots, m$ - est vérifiée.

Dans un premier temps, nous écrirons le problème "primal-dual" associé à (PI) sous la forme de la recherche d'un zéro d'un opérateur maximal monotone.

Ensuite, nous énoncerons l'algorithme "primal-dual" et nous étudierons sa convergence.

Enfin, nous vérifierons que par un choix particulier de fonction de Bregman, nous retrouvons la méthode proximale quadratique des multiplicateurs proposée par Rockafellar [17, paragraphe 5].

5.1 Formulation "primal-dual" du problème (PI)

Soit L *la fonction lagrangienne* associée au problème (PI). L est définie de $\mathbb{R}^n \times \mathbb{R}^m$ dans $\mathbb{R} \cup \{-\infty, +\infty\}$ par

$$L(x,p) = \begin{cases} f(x) + \sum_{i=1}^m p_i g_i(x) & \text{si } x \in X \quad p \in \overline{\Omega}^+ \\ -\infty & \text{si } x \in X \quad p \notin \overline{\Omega}^+ \\ +\infty & \text{si } x \notin X \end{cases}$$

Notons que cette fonction est *convexe - concave*, c'est-à-dire convexe par rapport à la première variable et concave par rapport à la seconde.
De plus la fonction L est propre, s.c.i en x pour tout y et s.c.s. en y pour tout x.

Nous savons que $\overline{x} \in \mathbb{R}^n$ et $\overline{p} \in \mathbb{R}^m$ sont respectivement solutions du primal (PI) et du dual (DI) si et seulement si $(\overline{x}, \overline{p})$ est un *point-selle* de la fonction L [19, théorème 23.8, p. 281], c'est-à-dire

$$L(\overline{x}, p) \leq L(\overline{x}, \overline{p}) \leq L(x, \overline{p}) \quad \forall\, x \in \mathbb{R}^n,\ \forall\, y \in \mathbb{R}^m\ .$$

Dès lors les problèmes (PI) et (DI) peuvent se réécrire sous la forme équivalente :

(S1) Trouver un point-selle de la fonction L dans $\mathbb{R}^n \times \mathbb{R}^m$.

De plus, en vertu de la caractérisation de la recherche d'un point-selle en terme du sous-différentiel [16, p. 40], le problème (S1) est équivalent à :

$$\text{(S2)} \qquad \text{Trouver } (\overline{x},\overline{p}) \in I\!R^n \times I\!R^m \text{ tel que } (0,0) \in \partial L(\overline{x},\overline{p})$$

$$\text{où } \partial L(\overline{x},\overline{p}) := \{(x^*,p^*) \in I\!R^n \times I\!R^m \ \text{ tel que } \ L(u,\overline{p}) \geq L(\overline{x},\overline{p}) + < x^*, u-\overline{x} > \ \forall\, u \in I\!R^n$$
$$\text{et} \quad L(\overline{x},v) \leq L(\overline{x},\overline{p}) + < p^*, v-\overline{p} > \ \forall\, v \in I\!R^m\}$$

c'est-à-dire $\partial L(\overline{x},\overline{p}) := \partial_x L(\overline{x},\overline{p}) \times \partial_p L(\overline{x},\overline{p})$.

Enfin, puisque à chaque fonction convexe - concave, propre, s.c.i. en x pour tout y, s.c.s. en y pour tout x, est associé un opérateur maximal monotone [19, corollaire 37.5.2, p. 396], nous concluons que le problème (S2) est équivalent à

$$\text{(S3)} \qquad \text{Trouver } (\overline{x},\overline{p}) \in I\!R^n \times I\!R^m \text{ tel que } (0,0) \in K(\overline{x},\overline{y})$$

où *l'opérateur maximal monotone* $K(x,p)$ est défini par

$$K(x,p) = \{(x^*,-p^*) \text{ tel que } (x^*,p^*) \in \partial L(x,p)\}\ . \tag{5.1}$$

5.2 Algorithme primal-dual

Soient $\{\lambda_k\}$ une suite de scalaires strictement positifs et h une fonction de Bregman de zone $S \subseteq I\!R^n \times I\!R^m$ telle que $D(K) \subseteq S$.

Considérons un couple $(x^0,p^0) \in S$ et applliquons l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman au problème (S3). Nous obtenons alors une suite $\{(x^k,p^k)\}_{k=0}^{\infty}$ telle que

$$(x^{k+1},p^{k+1}) = ((\nabla h + \lambda_k K)^{-1} \circ \nabla h)(x^k,p^k) \tag{5.2}$$

Dans le théorème suivant, nous adaptons l'itération de la suite $\{(x^k,p^k)\}$ au problème (PI).

Théorème 5.1 : *Forme de l'itération.*
Soient $\{\lambda_k\}$ une suite de scalaires strictement positifs; h_X une fonction de Bregman de zone $S_X \subseteq I\!R^n$ et h_p une fonction de Bregman de zone $S_p \subseteq I\!R^m$.
Nous supposons également $X \subseteq S_X, \overline{\Omega}^+ \subseteq S_p$ et $\Omega^+ \subseteq R(\nabla h_p)$.
Considérons la suite $\{(x^k, p^k)\}_{k=0}^{\infty}$ définie par

$$\begin{cases} x^{k+1} = \text{argmin}_{x \in I\!R^n} \{f_X(x) + \frac{1}{\lambda_k} h^{*+}(\nabla h_p(p^k) + \lambda_k g(x)) \\ \qquad\qquad\qquad\qquad + \frac{1}{\lambda_k} D h_X(x, x^k)\} \\ p^{k+1} = \nabla h_p^{*+}(\nabla h_p(p^k) + \lambda_k g(x^{k+1})) \, . \end{cases}$$

Alors
$$(x^{k+1}, p^{k+1}) = ((\nabla h + \lambda_k K)^{-1} \circ \nabla h)(x^k, p^k) \quad \text{pour tout } k \geq 0 \, ,$$
où h est la somme directe des fonctions h_X et h_p c'est-à-dire $h = h_X + h_p$.

Notons que la fonction h ainsi définie est une fonction de Bregman de zone $S = S_X \times S_p$ (cfr théorème 2.3).

De plus, en vertu de la proposition A.IV.6, on a

$$D(K) = D(\partial L) \subseteq \text{dom } L \subseteq X \times \overline{\Omega}^+$$

Or par hypothèse $X \times \overline{\Omega}^+ \subseteq S_X \times S_p$, ainsi nous obtenons que $D(K) \subseteq S$.

Avant de prouver ce théorème, évaluons le sous-différentiel de la fonction convexe - concave $L(x,p)$ par rapport aux variables x et p.

Lemme 5.1 :
Soient deux vecteurs x et p tels que $x \in X, g(x) \leq 0$ et $p \in \overline{\Omega}^+$.
Alors $\partial_p L(x,p) = \{g(x)\} - N_{\overline{\Omega}^+}(p)$.

Preuve du lemme 5.1 :

Evaluons d'abord $\partial_{p_i} L(x,p)$ où $i \in \{1, \cdots, m\}$.
Considérons la fonction C définie de $I\!R^+$ dans $I\!R$ par $C(p_i) = p_i g_i(x)$.
Il est évident que C est propre et concave. De plus $\partial_{p_i} L(x,p) = \partial C(p_i)$.

Or, par définition, on a

$$\begin{aligned} & \partial C(p_i) = \{y_i \in I\!R \text{ tel que } z_i g_i(x) \leq p_i g_i(x) + y_i(z_i - p_i) \ \forall\ z_i \geq 0\} \\ \Leftrightarrow \ & \partial C(p_i) = \{y_i \in I\!R \text{ tel que } (z_i - p_i) g_i(x) \leq y_i(z_i - p_i) \ \forall\ z_i \geq 0\} \ . \end{aligned}$$

Par conséquent, il résulte que

$$\partial p_i L(x,p) = \partial C(p_i) = \begin{cases} \{g_i(x)\} & \text{si } p_i > 0 \\ [g_i(x), \infty[& \text{si } p_i = 0 \ . \end{cases}$$

La thèse suit alors de la définition du cône normal $N_{\overline{\Omega}^+}$. ■

<u>**Lemme 5.2 :**</u>
Soient deux vecteurs x et p tels que $x \in X$ et $p \in \overline{\Omega}^+$.
Alors $\partial_x L(x,p) = \partial f(x) + N_X(x) + \sum_{i=1}^m \ p_i \partial g_i(x)$.

<u>**Preuve :**</u>

- Remarquons dès à présent

$$ri(\text{dom} \ \ f) \cap ri(\text{dom} \ \ g) \cap ri(\text{dom} \ \ \delta_X) \neq \emptyset \ . \tag{5.3}$$

  En effet, comme par hypothèse $X \subseteq$dom $f \cap$ dom $g \cap$ dom δ_X, il suit d'une propriété de l'intérieur relatif que (cfr proposition A.II.3)

$$ri(X) \subseteq ri(\text{dom} \ \ f) \cap ri(\text{dom} \ \ g) \cap ri(X) \ ;$$

  Or $ri(X) \neq \emptyset$ car X est convexe (cfr proposition A.II.4).

- Par conséquent, nous déduisons de la relation (5.3) et de la propriété de sommation du sous-différentiel (cfr proposition A.IV.12) que

$$\begin{aligned} & \partial_x L(x,p) = \partial(f(.) + \delta_X(.) + \sum_{i=1}^m p_i^{k+1} g(.))(x) \\ \Leftrightarrow \ & \partial_x L(x,p) = \partial f() + N_X(x) + \partial(\sum_{i=1}^m p_i g_i(.))(x) \ . \end{aligned}$$

  Or il suit de la proposition A.IV.15 que

$$\partial(\sum_{i=1}^m p_i g_i(.))(x) = \sum_{i=1}^m p_i \partial g_i(x) + N_{(\cap_{i=1}^m \text{dom } g_i)}(x) \ ,$$

  ainsi, nous obtenons que

$$\partial_x L(x,p) = \partial f(x) + N_X(x) + \sum_{i=1}^m p_i \partial g_i(x) + N_{(\cap_{i=1}^m \text{dom } g_i)}(x) \ .$$

- La thèse suit alors d'une propriété du cône normal. En effet, puisque $X \subseteq \cap_{i=1}^m$ dom g_i, la proposition A.V.3 nous assure que $N_X + N_{(\cap_{i=1}^m \text{dom } g_i)} = N_X$ ■

Prouvons maintenant le théorème relatif à la forme de l'itération de l'algorithme primal-dual.

<u>**Preuve du théorème 5.1 :**</u>

<u>Pas 1 :</u> Montrons d'abord que

$$\frac{1}{\lambda_k}(\nabla h_X(x^k) - \nabla h_X(x^{k+1})) \in \partial_x L(x^{k+1}, p^{k+1}) \;. \tag{5.4}$$

Définissons les deux opérateurs suivants :

$$\forall\; u \in I\!R^m \quad P_k(u) = \frac{1}{\lambda_k} h^{*+}(\nabla h_p(p^k) + \lambda_k u)$$

$$\forall\; x \in I\!R^n \quad G_k(x) = \begin{cases} P_k(g(x)) & \text{si } g(x) \in I\!R^m \\ +\infty & \text{si } g_i(x) = +\infty \;\forall\; i = 1, \cdots, m \end{cases}$$

- Il suit des définitions de G_k et f_X que

$$x^{k+1} = \text{argmin}_{\,x \in I\!R^n} \; \{f_X(x) + \frac{1}{\lambda_k} h_p^{*+}(p^{(p^k)}) + \lambda_k g(x)) + \frac{1}{\lambda_k} D_{h_X}(x, x^k)\}$$

$$\Leftrightarrow \quad x^{k+1} = \text{argmin}_{\,x \in I\!R^n} \; \{f(x) + \delta_X(x) + G_k(x) + \frac{1}{\lambda_k} D_{h_X}(x, x^k)\}$$

Ainsi, écrivant la condition d'optimalité d'un problème de minimisation convexe, nous obtenons que

$$0 \in \partial\{f(.) + \delta_X(.) + G_k(.) + \frac{1}{\lambda_k} D_{h_X}(., x^k)\}(x^{k+1}) \tag{5.5}$$

De plus, par hypothèse, $X \subseteq$dom $f, X \subseteq S_X$ et $X \subseteq$dom $g =$dom G_k car dom $P_k = I\!R^m$ (cfr lemme 4.1). Dès lors il suit de la proposition A.II.8 que

$$ri(X) \subseteq ri(\text{dom } f) \cap ri(\text{dom } G_k) \cap ri(\text{dom } D_{h_X}(., x^k)) \;.$$

Or $ri(X) \neq \emptyset$ car X est convexe, d'où

$$ri(\text{dom } f) \cap ri(X) \cap ri(\text{dom } G_k) \cap ri(\text{dom } D_{h_X}(., x^k)) \neq \emptyset$$

Par conséquent, la propriété de sommation du sous-différentiel nous assure que

$$0 \in \partial f(x^{k+1}) + N_X(x^{k+1}) + \partial G_K(x^{k+1}) + \frac{1}{\lambda_k} \partial_1 (D_{h_X}(., x^k))(x^{k+1}) \tag{5.6}$$

où ∂_1 désigne l'ensemble des sous-gradients par rapport à la première coordonnée.

- Evaluons maintenant $\partial G_K(x^{k+1})$.

D'une part l'opérateur P_k est différentiable et croissant sur $I\!R^m$ (cfr lemme 4.4). D'autre part les fonctions $g_i \quad i = 1, \cdots, m$ sont propres, convexes et $\cap_{i=1}^m ri(\text{dom } \ g_i) \neq \emptyset$ car $ri(X) \subseteq \cap_{i=1}^m ri(\text{dom } \ g_i)$.

Dès lors il suit de la proposition 4.17 que

$$\begin{aligned} &\partial G_k(x^{k+1}) = \sum_{i=1}^m [\nabla P_k(g(x^{k+1}))]_i \partial g_i(x^{k+1}) + N_{(\cap_{i=1}^m \text{dom } g_i)}(x^{k+1}) \\ \Leftrightarrow \quad &\partial G_k(x^{k+1}) = \sum_{i=1}^m [\nabla h_p^{*+}[\nabla h_p(p^k) + \lambda_k g(x)]]_i \partial g_i(x^{k+1}) \\ &\qquad + N_{(\cap_{i=1}^m \text{dom } g_i)}(x^{k+1}) \\ \Leftrightarrow \quad &\partial G_k(x^{k+1}) = \sum_{i=1}^m p_i^{k+1} \partial g_i(x^{k+1}) + N_{(\cap_{i=1}^m \text{dom } g_i)}(x^{k+1}) \end{aligned} \tag{5.7}$$

- Ainsi, remplaçant $\partial(G_k)$ par sa valeur (cfr (5.7)) dans la relation (5.6), nous obtenons que

$$0 \in \partial f(x^{k+1}) + N_X(x^{k+1}) + \sum_{i=1}^m p_i^{k+1} \partial g_i(x^{k+1}) + N_{(\cap_{i=1}^m \text{dom } g_i)}(x^{k+1}) + \frac{1}{\lambda_k} \partial_1(D_{h_X}(., x^k))(x^{k+1}) . \tag{5.8}$$

Or par définition $D_{h_X}(x, y) = h_X(x) - h_X(y) - < \nabla h_X(y), x - y >$, d'où

$$\partial_1(D_{h_X}(., x^k))(x^{k+1}) = \frac{1}{\lambda_k}(\nabla h_X(x^{k+1}) - \nabla h_X(x^k)) \tag{5.9}$$

De plus puisque $X \subseteq \cap_{i=1}^m$ dom g_i, il suit d'une propriété du cône normal (cfr proposition A.V.3) que

$$N_X + N_{(\cap_{i=1}^m \text{dom } g_i)} = N_X . \tag{5.10}$$

Par conséquent, il résulte des relations (5.8), (5.9) et (5.10) que

$$0 \in \partial f(x^{k+1}) + N_X(x^{k+1}) + \sum_{i=1}^m p_i^{k+1} \partial g_i(x^{k+1}) + \frac{1}{\lambda_k}(\nabla h_X(x^{k+1}) - \nabla h_X(x^k)) ,$$

c'est-à-dire

$$\frac{1}{\lambda_k}(\nabla h_X(x^k) - \nabla h_X(x^{k+1})) \in \partial f(x^{k+1}) + N_X(x^{k+1}) + \sum_{i=1}^m p_i^{k+1} \partial g_i(x^{k+1})$$

La thèse suit alors de la définition de $\partial_X l(x^{k+1}, p^{k+1})$ (cfr lemme 5.2).

Pas 2 : Vérifions ensuite que

$$\frac{1}{\lambda_k}(\nabla h_p(p^k) - \nabla h_p(p^{k+1})) \in -\partial_p l(x^{k+1}, p^{k+1}) . \tag{5.11}$$

Comme la fonction h_p^{*+} est propre, convexe et fermée (cfr proposition 4.12), la proposition 4.2 nous assure que ∇h_p^{*+} et $\partial(h_p^{*+})^*$ sont des opérateurs inverses.
Ainsi, nous obtenons que

$$\begin{aligned} & p^{k+1} = \nabla h_p^{*+}(\nabla h_p(p^k) + \lambda_k g(x^{k+1})) \\ \Leftrightarrow \quad & \nabla h_p(p^k) + \lambda_k g(x^{k+1}) \in \partial(h_p^{*+})^*(p^{k+1}) \, . \end{aligned} \tag{5.12}$$

Or la fonction $h_p^{*+} = h_p + \delta^+$ est fermée (cfr proposition 4.10), d'où $(h_p^{*+})^{**} = (h_p + \delta^+)^{**} = h_p + \delta^+$ (cfr proposition A.VI.4).
Dès lors la relation (5.12) se réécrit

$$\nabla h_p(p^k) + \lambda_k g(x^{k+1}) \in \partial(h_p + \delta^+)(p^{k+1}) \, . \tag{5.13}$$

D'autre part $ri(S_p) \cap ri(\overline{\Omega}^+) \neq \emptyset$. En effet, comme par hypothèse $\Omega^+ \subseteq S_p$, il suit d'une propriété de l'intérieur relatif (cfr proposition A.II.8) que $\Omega^+ \subseteq ri(S_p)$.
Par conséquent $\partial(h_p + \delta^+)(p^{k+1}) = \nabla h_p(p^{k+1}) + \partial\delta^+(p^{k+1})$ (cfr proposition A.IV.12).

Ainsi la relation (5.13) est équivalente à

$$\nabla h_p(p^k) + \lambda_k g(x^{k+1}) \in \nabla h_p(p^{k+1}) + \partial\delta^+(p^{k+1}) \, ,$$

ou encore

$$\frac{1}{\lambda_k}\{\nabla h_p(p^k) - \nabla h_p(p^{k+1})\} \in -\{\{g(x^{k+1})\} - N_{\overline{\Omega}^+}(p^{k+1})\} \, .$$

La thèse suit alors de la définition de $\partial_p l(x^{k+1}, p^{k+1})$ (cfr lemme 5.1).
<u>Pas 3 :</u> Prouvons que

$$(x^{k+1}, p^{k+1}) = ((\nabla h + \lambda_k K)^{-1} \circ \nabla h)(x^k, p^k) \quad \forall\, k \geq 0$$

Il résulte des relations (5.4) et (5.11) que

$$\begin{cases} \nabla h_X(x^k) \in \lambda_k \partial_x l(x^{k+1}, p^{k+1}) + \nabla h_X(x^{k+1}) \\ \nabla h_p(p^k) \in -\lambda_k \partial_p l(x^{k+1}, p^{k+1}) + \nabla h_p(p^{k+1}) \end{cases}$$

$$\begin{aligned} & \Leftrightarrow (\nabla h_X(x^k), \nabla h_p(p^k)) \in \lambda_k K(x^{k+1}, p^{k+1}) + (\nabla h_X(x^{k+1}), \nabla h_p(p^{k+1})) \\ & \Leftrightarrow \nabla h(x^k, p^k) \in (\lambda_k K + \nabla h)(x^{k+1}, p^{k+1}) \\ & \Leftrightarrow (x^{k+1}, p^{k+1}) \in ((\nabla h + \lambda_k K)^{-1} \circ \nabla h)(x^k, p^k) \, . \end{aligned}$$

Or l'opérateur $(\nabla h + \lambda_k K)^{-1} \circ \nabla h$ est univoque car K est un opérateur monotone (cfr proposition 2.7). Ainsi, nous concluons que

$$(x^{k+1}, p^{k+1}) = ((\nabla h + \lambda_k K)^{-1} \circ \nabla h)(x^k, p^k) \, .$$

Ce qu'il fallait démontrer. ■

Il est maintenant temps d'énoncer l'algorithme primal-dual.

Algorithme 5.1 :

Soient $\{\lambda_k\}$ une suite de scalaires strictement positifs, h_X une fonction de Bregman de zone $S_X \subseteq I\!R^n$ telle que $X \subseteq S_X$ et h_p une fonction de Bregman de zone $S_p \subseteq I\!R^m$ telle que $\overline{\Omega}^+ \subseteq S_p$ et $\Omega^+ \subseteq R(\nabla h_p)$.
Etant donné $(x^0, p^0) \in X \times \overline{\Omega}^+$, l'algorithme de la *méthode proximale des multiplicateurs utilisant des fonctions de Bregman* génère une suite $\{(x^k, p^k)\}_{k=0}^{\infty}$ telle que

$$\begin{cases} x^{k+1} = \operatorname{argmin}_{x \in I\!R^n} \{f_X(x) + \frac{1}{\lambda_k} h_p^{*+}(\nabla h_p(p^k) + \lambda_k g(x)) + \frac{1}{\lambda_k} D_{h_X}(x, x^k)\} \\ p^{k+1} = \nabla h_p^{*+}(\nabla h_p(p^k) + \lambda_k g(x^{k+1})) \, . \end{cases}$$

Comparant cet algorithme avec celui de [17, paragraphe 5], nous remarquons que les termes linéaires et quadratiques du lagrangien augmenté ont été remplacés par un terme de pénalité plus général. De plus le terme proximal quadratique a été remplacé par une fonction "distance" D_h.

5.3 Conditions d'existence et d'unicité des itérés

Théorème 5.2 :

Soient $\{\lambda_k\}$ une suite de scalaires strictement positifs, h_X une fonction de Bregman de zone $S_X \subseteq I\!R^n$ telle que $X \subseteq S_X$ et h_p une fonction de Bregman de zone $S_p \subseteq I\!R^m$ telle que $\overline{\Omega}^+ \subseteq S_p$ et $\Omega^+ \subseteq R(\nabla h_p)$.
Supposons également $R(\nabla h_X) = I\!R^n$ et $R(\nabla h_p) = I\!R^m$.
Alors la suite $\{(x^k, p^k)\}$ générée par la méthode proximale des multiplicateurs utilisant des fonctions de Bregman, c'est-à-dire

$$\begin{cases} x^{k+1} = \text{argmin}_{\,x \in I\!R^n} \; \{f_X(x) + \frac{1}{\lambda_k} h_p^{*+}(\nabla h_p(p^k) + \lambda_k g(x)) + \frac{1}{\lambda_k} D_{h_X}(x, x^k)\} \\ p^{k+1} = \nabla h_p^{*+}(\nabla h_p(p^k) + \lambda_k g(x^{k+1})) \, . \end{cases}$$

est bien définie quel que soit le point de départ $(x^0, p^0) \in I\!R^n \times \overline{\Omega}^+$.

Preuve :

Par le théorème 5.1, nous savons que

$$(x^{k+1}, p^{k+1}) = ((\nabla h + \lambda_k K)^{-1} \circ \nabla h)(x^k, p^k)$$

où la fonction h est la somme directe des fonctions h_X et h_p.

De plus il est évident que $R(\nabla h) = I\!R^{n+m}$.
En effet $R(\nabla h_X) = I\!R^n$ et $R(\nabla h_p) = I\!R^m$.

La thèse suit alors du théorème 2.6 appliqué à l'opérateur maximal monotone K. ∎

5.4 Convergence vers une solution des problèmes (PI) et (DI)

Nous appelons *couple solution du problème* (PI) tout couple $(\overline{x}, \overline{p})$ tel que $\overline{x}$ est une solution de (PI) et $\overline{y}$ une solution de (DI).

Théorème 5.3 :
Soient $\{\lambda_k\}$ une suite de scalaires strictement positifs, h_X une fonction de Bregman de zone $S_X \subseteq I\!R^n$ telle que $X \subseteq S_X$ et h_p une fonction de Bregman de zone $S_p \subseteq I\!R^m$ telle que $\overline{\Omega}^+ \subseteq S_p$ et $\Omega^+ \subseteq R(\nabla h_p)$.
Supposons également que $\inf_{k\geq 0} \{\lambda_k\} > 0$.
Alors la suite $\{(x^k, p^k)\}$ générée par la méthode proximale des multiplicateurs utilisant des fonctions de Bregman, c'est-à-dire

$$\begin{cases} x^{k+1} = \text{argmin}_{x \in I\!R^n} \{f_X(x) + \frac{1}{\lambda_k} h_p^{*+}(\nabla h_p(p^k) + \lambda_k g(x)) + \frac{1}{\lambda_k} D_{h_X}(x, x^k)\} \\ p^{k+1} = \nabla h_p^{*+}(\nabla h_p(p^k) + \lambda_k g(x^{k+1})) \end{cases}$$

converge vers un couple solution du problème (PI) s'il en existe une. Dans le cas contraire, au moins une des deux suites $\{x^k\}$ ou $\{p^k\}$ est non bornée.

Preuve :

Par le théorème 5.1, nous savons que

$$(x^{k+1}, p^{k+1}) = ((\nabla h + \lambda_k K)^{-1} \circ \nabla h)(x^k, p^k)$$

où la fonction h est la somme directe des fonctions h_X et h_p.

De plus, il suit de la proposition A.IV.6 que

$$\overline{D(K)} = \overline{D(\partial L)} \subseteq \overline{\text{dom } L} \subseteq S_X \times S_p \, .$$

Ainsi, nous obtenons que $\overline{D(K)} \subseteq S = S_X \times S_p$.

La thèse suit alors du théorème 2.7 appliqué à l'opérateur maximal monotone K. ∎

5.5 Cas particulier : $h_X(x) = \frac{1}{2}||x||^2$ et $h_p(p) = \frac{1}{2}||p||^2$

Considérons les fonctions de Bregman h_X et h_p définies respectivement sur $I\!R^n$ et $I\!R^m$ par $h_X(x) = \frac{1}{2}||x||^2$ et $h_p(p) = \frac{1}{2}||p||^2$.

Remarquons dès à présent que $X \subseteq S_X$ et $\overline{\Omega}^+ \subseteq S_p$ car $S_X = I\!R^n$ et $S_p = I\!R^m$. De plus $R(\nabla h_X) = I\!R^n$ et $R(\nabla h_p) = I\!R^m$ (exemple 2.11).

D'autre part, comme la fonction h_p est séparable, nous déduisons des propositions 4.7 et 4.8 que $h_p^{*+}(p) = h_p^*(\underline{\max}\{p, \nabla h_p(0)\})$ et $\nabla h_{p_i}^{*+}(p_i) = \max\{0, \nabla h_{p_i}(0)\}$ où $\underline{\max}(x,y)$ désigne le vecteur dont chaque composante est le maximum entre x_i et y_i.
Or $h_p^*(p) = \frac{1}{2}||p||^2$, $\nabla h_p(p) = p$ et $D_{h_X}(x,y) = \frac{1}{2}||x-y||^2$.

Par conséquent, pour tout $(x^0, p^0) \in X \times \overline{\Omega}^+$, la méhode proximale des multiplicateurs utilisant les fonctions de Bregman $h_X = \frac{1}{2}||x||^2$ et $h_p = \frac{1}{2}||p||^2$ génère une suite $\{(x^k, p^k)\}$ telle que :

$$\begin{cases} x^{k+1} = \text{argmin}_{x \in I\!R^n} \{ f_X(x) + \frac{1}{\lambda_k} h_p^{*+}(\nabla h_p(p^k) + \lambda_k g(x)) \\ \qquad\qquad + \frac{1}{\lambda_k} Dh_X(x, x^k)\} \\ p^{k+1} = \nabla h_p^{*+}(\nabla h_p(p^k) + \lambda_k g(x^{k+1})) \end{cases}$$

$$\Leftrightarrow \begin{cases} x^{k+1} = \text{argmin}_{x \in I\!R^n} \{ f_X(x) + \frac{1}{2\lambda_k} \sum_{i=1}^m (\max(0, p_i^k + \lambda_k g_i(x)))^2 \\ \qquad\qquad + \frac{1}{2\lambda_k} \sum_{i=1}^n (x_i - x_i^k)^2 \} \\ p^{k+1} = \underline{\max}\{0, p_k + \lambda_k g(x^{k+1})\} \end{cases}$$

Comme annoncé précédemment, il s'agit de *la méthode proximale des multiplicateurs proposée par Rockafellar* [17, paragraphe 5].

Chapitre 6

Première approche de la régularisation non linéaire d'une somme d'opérateurs maximaux monotones

Dans ce chapitre, nous désirons appliquer l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman (cfr algorithme 2.1) au problème suivant :

$$\text{(P)} \qquad \text{trouver } x \in I\!R^n \text{ tel que } 0 \in (A+B)(x) \; .$$

où A et B sont deux opérateurs maximaux monotones de domaine respectif $D(A) \subseteq I\!R^n$ et $D(B) \subseteq I\!R^n$.

Cependant, cela ne peut se faire car la somme d'opérateurs maximaux monotones n'est pas nécessairement maximale monotone. Nous allons donc *régulariser* le problème (P). Plus précisément, nous allons approximer l'un des deux opérateurs A ou B (s.p.d.g. $A' \simeq A$) de sorte que l'opérateur $A' + B$ soit maximal monotone.

Nous commencerons ce chapitre par un bref rappel d'une technique de régularisation classique proposée par P. Mahey et Pham Dinh Tao [14].

Ensuite, nous présenterons les difficultés rencontrées lors de la généralisation de cette technique à l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman.

6.1 Technique de régularisation classique

Dans le cas classique, nous remplaçons l'un des deux opérateurs maximaux monotones A ou B (par exemple A) par son *approximation de Moreau Yosida* :

$$A_\lambda := \frac{I - (I + \lambda A)^{-1}}{\lambda} \text{ avec } \lambda > 0 \ .$$

Puisque l'opérateur A_λ est monotone et lipschitzien [2, proposition 2.6, p. 28], nous déduisons de la proposition 2.5 que l'opérateur $A_\lambda + B$ est maximal monotone. Ainsi, nous pouvons appliquer la théorie de Rockafellar (cfr paragraphe 2.3) au *problème régularisé*

$$(\mathrm{P})_\lambda \qquad \text{trouver } x_\lambda \in I\!R^n \text{ tel que } 0 \in (A_\lambda + B)(x_\lambda) \ .$$

Le problème (P_λ) est donc équivalent à la recherche d'un point fixe de l'opérateur $(I + c(A_\lambda + B))^{-1}$ où c et λ sont deux scalaires strictement positifs.

D'autre part tout point fixe de $(I + c(A_\lambda + B))^{-1}$ est un point fixe de l'opérateur $(I + \lambda B)^{-1}(I + \lambda A)^{-1}$.
En effet,

$$\begin{aligned}
& (I + c(A_\lambda + B))^{-1}(x_\lambda) = x_\lambda && \text{avec } c > 0 \\
\Leftrightarrow \quad & x_\lambda \in (I + c(\frac{I - (I + \lambda A)^{-1}}{\lambda} + B))(x_\lambda) && \text{avec } c > 0 \\
\Leftrightarrow \quad & x_\lambda \in x_\lambda + \frac{c}{\lambda} x_\lambda - \frac{c}{\lambda}(I + \lambda A)^{-1}(x_\lambda) + cB(x_\lambda) && \text{avec } c > 0 \\
\Leftrightarrow \quad & 0 \in (I + \lambda B)(x_\lambda) - (I + \lambda A)^{-1}(x_\lambda) \\
\Leftrightarrow \quad & (I + \lambda A)^{-1}(x_\lambda) \in (I + \lambda B)(x_\lambda) \\
\Leftrightarrow \quad & (I + \lambda B)^{-1}(I + \lambda A)^{-1}(x_\lambda) = x_\lambda
\end{aligned}$$

En conclusion, résoudre le problème (P_λ) revient à trouver un point fixe de l'opérateur $(I + \lambda B)^{-1}(I + \lambda A)^{-1}$.
D'où l'idée d'écrire l'itération du point fixe

$$x^{t+1} = (I + \lambda B)^{-1}(I + \lambda A)^{-1}(x^t) \ . \tag{6.1}$$

Afin de mettre en évidence la dépendance vis à vis du paramètre de régularisation λ, nous écrivons l'itération doublement indicée

$$x_k^{t+1} = (I + \lambda_k B)^{-1}(I + \lambda_k A)^{-1}(x_k^t) \tag{6.2}$$

où $\{\lambda_k\}$ est une suite de nombres réels positifs.

Il s'agit donc d'une méthode en deux temps. D'abord, nous itérons sur l'indice t pour un k fixé. Ensuite nous augmentons k en générant ainsi un nouveau cycle d'itérations.

Puisque les opérateurs $(I + \lambda B)^{-1}$ et $(I + \lambda A)^{-1}$ sont des *pseudo-contractions* [22, p. 24], l'opérateur $(I + \lambda B)^{-1}(I + \lambda A)^{-1}$ résultant de leur composition est une contraction. Dès lors pour λ_k fixé, la suite générée par l'itération (6.2) converge vers une solution x_k du problème régularisé correspondant (P_{λ_k}), si une telle solution existe (cfr [22, p. 25]). Ensuite si nous faisons tendre la suite de paramètres de régularisation $\{\lambda_k\}$ vers zéro, les solutions des problèmes régularisés convergent vers une solution du problème général (P).

6.2 Technique de régularisation non linéaire

6.2.1 Forme de l'itération

Soient h une fonction de Bregman de zone S telle que $D(A) \subseteq S, D(B) \subseteq S$ et $\{\lambda_k\}_{k=0}^{\infty}$ une suite de scalaires strictement positifs. Par analogie avec le cas classique, nous souhaitons que la suite définie par l'itération

$$x_k^{t+1} = ((\nabla h + \lambda_k B)^{-1} \circ \nabla h)((\nabla h + \lambda_k A)^{-1} \circ \nabla h)(x_k^t)$$

converge vers un zéro de l'opérateur $A + B$ lorsque nous faisons tendre t vers l'infini et la suite $\{\lambda_k\}$ vers zéro.

Par conséquent, nous remplaçons l'opérateur A par son approximation :

$$A_{h\lambda} := \frac{\nabla h - \nabla h(\nabla h + \lambda A)^{-1}\nabla h}{\lambda} \text{ où } \lambda > 0$$

Ainsi, le *problème régularisé* s'énonce :

$$(P_{h\lambda}) \qquad \text{trouver } x_{h\lambda} \in I\!R^n \text{ tel que } 0 \in (A_{h\lambda} + B)(x_{h\lambda})$$

Vérifions maintenant que l'opérateur $A_{h\lambda}$ est *monotone et lipschitzien.*

Proposition 6.1 :
Soient A un opérateur maximal monotone et h une fonction de Bregman de zone S telle que $D(A) \subseteq S$.
Alors l'opérateur $A_{h\lambda}$ défini par

$$A_{h\lambda} := \frac{\nabla h - \nabla h(\nabla h + \lambda A)^{-1} \nabla h}{\lambda} \text{ où } \lambda > 0$$

est monotone.

Preuve :

Considérons l'opérateur P_λ défini sur $D(A)$ par $P_\lambda := (\nabla h + \lambda A)^{-1} \circ \nabla h$.

- Il est tout d'abord immédiat que $\forall\, x_1, x_2 \in D(A) = D(A_{h\lambda})$

$$\begin{aligned} < A_{h\lambda}x_1 - A_{h\lambda}x_2, x_1 - x_2 > &= < A_{h\lambda}x_1 - A_{h\lambda}x_2, P_\lambda x_1 - P_\lambda x_2 > \\ &+ < A_{h\lambda}x_1 - A_{h\lambda}x_2, (I - P_\lambda)x_1 - (I - P_\lambda)x_2 > \end{aligned} \tag{6.3}$$

- D'une part vérifions que

$$< A_{h\lambda}x_1 - A_{h\lambda}x_2, P_\lambda x_1 - P_\lambda x_2 > \geq 0 \,. \tag{6.4}$$

Il suit de la définition des opérateurs $A_{h\lambda}$ et P_λ que

$$\begin{aligned} & P_\lambda x_1 = ((\nabla h + \lambda A)^{-1} \circ \nabla h) x_1 \\ \Leftrightarrow\ & \nabla h(x_1) - \nabla h(P_\lambda x_1) \in \lambda A(P_\lambda x_1) \\ \Leftrightarrow\ & A_{h\lambda}(x_1) \in A(P_\lambda x_1) \,. \end{aligned}$$

De même, nous obtenons que $A_{h\lambda}(x_2) \in A(P_\lambda x_2)$.
La thèse résulte alors de la monotonie de l'opérateur A (cfr définition 2.7).

- D'autre part, montrons que

$$< A_{h\lambda}x_1 - A_{h\lambda}x_2, (I - P_\lambda)x_1 - (I - P_\lambda)x_2 > \geq 0 \,. \tag{6.5}$$

Définissons $z_1 = (I - P_\lambda)x_1$ et $z_2 = (I - P_\lambda)x_2$.
Ainsi, comme $A_{h\lambda} = \frac{1}{\lambda}(\nabla h \circ (I - P_\lambda))$, nous obtenons que

$$< A_{h\lambda}x_1 - A_{h\lambda}x_2, (I - P_\lambda)x_1 - (I - P_\lambda)x_2 >= \frac{1}{\lambda} < \nabla h(z_1) - \nabla h(z_2), z_1 - z_2 > \ .$$

De plus le membre de droite de cette égalité est positif puisque l'opérateur ∇h est monotone.

- Par conséquent, nous déduisons des relations (6.3), (6.4) et (6.5) que

$$\forall\ x_1, x_2 \in D(A) = D(A_{h\lambda}) \quad < A_{h\lambda}x_1 - A_{h\lambda}x_2, x_1 - x_2 > \geq 0$$

Ce qu'il fallait démontrer. ∎

Dans le but de prouver que l'opérateur $A_{h\lambda}$ est lipschitzien, nous supposons que *le gradient de la fonction de Bregman h est un opérateur fortement monotone et lipschitzien.* Rappelons tout d'abord ces deux définitions.

Définition 6.1 :
L'opérateur A est *fortement monotone* de constante α s'il existe un scalaire $\alpha > 0$ tel que pour tout $z_1, z_2 \in D(A)$

$$< Az_1 - Az_2, z_1 - z_2 > \geq \alpha||z_1 - z_2||^2 \ .$$

Définition 6.2 :
L'opérateur A est *lipschitzien* de constante μ s'il existe un scalaire $\mu > 0$ tel que pour tout $z_1, z_2 \in D(A)$

$$||Az_1 - Az_2|| \leq \mu||z_1 - z_2||$$

Enonçons maintenant la proposition suivante :

Proposition 6.2 :
Soient A un opérateur maximal monotone et h une fonction de Bregman de zone $S \supseteq D(A)$. Supposons que ∇h est un opérateur fortement monotone de constante α et lipschitzien de constante μ.

Alors l'opérateur $A_{h\lambda}$ défini par

$$A_{h\lambda} := \frac{\nabla h - \nabla h(\nabla h + \lambda A)^{-1}\nabla h}{\lambda} \text{ avec } \lambda > 0$$

est lipschitzien.

Preuve :

Soient x_1, x_2 deux vecteurs distincts de $D(A)$ et l'opérateur P_λ défini sur $D(A)$ par $P_\lambda := (\nabla h + \lambda A)^{-1} \circ \nabla h$.
Il suit de l'inégalité de Cauchy Schwartz que

$$||A_{h\lambda}x_1 - A_{h\lambda}x_2||\ ||x_1 - x_2|| \geq < A_{h\lambda}x_1 - A_{h\lambda}x_2, x_1 - x_2 > ,$$

ou de façon équivalente :

$$\begin{aligned} ||A_{h\lambda}x_1 - A_{h\lambda}x_2||\ ||x_1 - x_2|| &\geq < A_{h\lambda}x_1 - A_{h\lambda}x_2, P_\lambda x_1 - P_\lambda x_2 > \\ &+ < A_{h\lambda}x_1 - A_{h\lambda}x_2, (I - P_\lambda)x_1 - (I - P_\lambda)x_2 > \quad (6.6) \end{aligned}$$

- Montrons d'abord que

$$< A_{h\lambda}x_1 - A_{h\lambda}x_2, P_\lambda x_1 - P_\lambda x_2 > \geq 0 . \quad (6.7)$$

Il suit de la définition de P_λ que

$$\begin{aligned} & P_\lambda x_1 = ((\nabla h + \lambda A)^{-1} \circ \nabla h)(x_1) \\ \Leftrightarrow\ & \nabla h(x_1) - \nabla h(P_\lambda x_1) \in \lambda A(P_\lambda x_1) \\ \Leftrightarrow\ & A_{h\lambda}(x_1) \in A(P_\lambda x_1) \end{aligned}$$

De même nous obtenons que $A_{h\lambda}(x_2) \in A(P_\lambda x_2)$.
La thèse suit alors de la monotonie de l'opérateur A.

- Montrons ensuite que

$$< A_{h\lambda}x_1 - A_{h\lambda}x_2, (I - P_\lambda)x_1 - (I - P_\lambda)x_2 > \geq \frac{\alpha\lambda}{\mu^2}||A_{h\lambda}x_1 - A_{h\lambda}x_2||^2 . \quad (6.8)$$

Définissant $z_1 = (I - P_\lambda)x_1$ et $z_2 = (I - P_\lambda)x_2$, nous déduisons de la définition de $A_{h\lambda}$ que

$$< A_{h\lambda}x_1 - A_{h\lambda}x_2, (I - P_\lambda)x_1 - (I - P_\lambda)x_2 >= \frac{1}{\lambda} < \nabla h z_1 - \nabla h z_2, z_1 - z_2 > .$$

Or l'opérateur ∇h est fortement monotone de constante α, d'où

$$< A_{h\lambda}x_1 - A_{h\lambda}x_2, (I - P_\lambda)x_1 - (I - P_\lambda)x_2 > \geq \frac{1}{\lambda}\alpha||z_1 - z_2||^2 ,$$

c'est-à-dire

$$< A_{h\lambda}x_1 - A_{h\lambda}x_2, (I - P_\lambda)x_1 - (x - P_\lambda)x_2 > \geq \frac{1}{\lambda}\alpha||(I - P_\lambda)x_1 - (I - P_\lambda)x_2||^2 . \quad (6.9)$$

D'autre part

$$||(I - P_\lambda)x_1 - (I - P_\lambda)x_2||^2 \geq \frac{\lambda^2}{\mu^2}||A_{h\lambda}x_1 - A_{h\lambda}x_2||^2 . \quad (6.10)$$

En effet, il résulte de la définition de $A_{h\lambda}$ que

$$||A_{h\lambda}x_1 - A_{h\lambda}x_2||^2 = \frac{1}{\lambda^2}||\nabla h(I - P_\lambda)x_1 - \nabla h(I - P_\lambda)x_2||^2 .$$

Or l'opérateur ∇h est lipschitzien de constante μ, par conséquent il suit que

$$||A_{h\lambda}x_1 - A_{h\lambda}x_2||^2 \leq \frac{\mu^2}{\lambda^2}||(I - P_\lambda)x_1 - (I - P_\lambda)x_2||^2$$

Ainsi regroupant les relations (6.9) et (6.10), nous obtenons que

$$< A_{h\lambda}x_1 - A_{h\lambda}x_2, (I - P_\lambda)x_1 - (I - P_\lambda)x_2 > \geq \frac{\alpha\lambda}{\mu^2}||A_{h\lambda}x_1 - A_{h\lambda}x_2||^2 .$$

- En conclusion, il résulte des inégalités (6.6), (6.7) et (6.8) que

$$||A_{h\lambda}x_1 - A_{h\lambda}x_2|| \; ||x_1 - x_2|| \geq \frac{\alpha\lambda}{\mu^2}||A_{h\lambda}x_1 - A_{h\lambda}x_2||^2$$

autrement dit,

$$||A_{h\lambda}x_1 - A_{h\lambda}x_2|| \leq \gamma||x_1 - x_2|| \text{ où } \gamma := \frac{\alpha\lambda}{\mu^2} > 0 .$$

Ce qu'il fallait démontrer. ∎

Comme *la somme d'un opérateur monotone, lipschitzien A_λ et d'un opérateur monotone B est maximale monotone* (cfr proposition 2.5), nous pouvons écrire l'itération du point fixe

$$((\nabla h + c(A_{h\lambda} + B))^{-1} \circ \nabla h)(x^t) = x^t \text{ où } c > 0$$

pour résoudre le problème régularisé $(P_{h\lambda})$.

De plus tout point fixe de l'opérateur $(\nabla h + c(A_{h\lambda} + B))^{-1} \circ \nabla h$ est un point fixe de l'opérateur $((\nabla h + \lambda B)^{-1} \circ \nabla h)((\nabla h + \lambda A)^{-1} \circ \nabla h)$.
En effet, après simplification, nous obtenons que

$$\begin{aligned}
& ((\nabla h + c(A_{h\lambda} + B))^{-1} \circ \nabla h)(x_{h\lambda}) = x_{h\lambda} && \text{avec } c > 0 \\
\Leftrightarrow\ & \nabla h(x_{h\lambda}) \in (\nabla h + c(A_{h\lambda} + B))(x_{h\lambda}) && \text{avec } c > 0 \\
\Leftrightarrow\ & \nabla h(x_{h\lambda}) \in \nabla h(x_{h\lambda}) + c(\frac{\nabla h - \nabla h(\nabla h + \lambda A)^{-1}\nabla h}{\lambda} + B)(x_{h\lambda}) && \text{avec } c > 0 \\
\Leftrightarrow\ & 0 \in \frac{c}{\lambda}\nabla h(x_{h\lambda}) + cB(x_{h\lambda}) - \frac{c}{\lambda}(\nabla h(\nabla h + \lambda A)^{-1}\nabla h)(x_{h\lambda}) && \text{avec } c > 0 \\
\Leftrightarrow\ & (\nabla h \circ (\nabla h + \lambda A)^{-1} \circ \nabla h)(x_{h\lambda}) \in (\nabla h + \lambda B)(x_{h\lambda}) \\
\Leftrightarrow\ & ((\nabla h + \lambda B)^{-1} \circ \nabla h)((\nabla h + \lambda A)^{-1} \circ \nabla h)(x_{h\lambda}) = x_{h\lambda}
\end{aligned}$$

Notons que la dernière égalité suit du caractère univoque des deux opérateurs $(\nabla h + \lambda B)^{-1} \circ \nabla h$ et $(\nabla h + \lambda A)^{-1} \circ \nabla h$ (cfr proposition 2.6).

En conclusion, *résoudre le problème $(P_{h\lambda})$ revient à trouver un point fixe de l'opérateur* $((\nabla h + \lambda A)^{-1} \circ \nabla h)((\nabla h + \lambda A)^{-1} \circ \nabla h)$. Nous pouvons donc écrire l'itération du point fixe :

$$x^{t+1} = ((\nabla h + \lambda B)^{-1} \circ \nabla h)((\nabla h + \lambda A)^{-1} \circ \nabla h)(x^t)\,. \tag{6.11}$$

Comme dans le cas classique, nous écrivons l'itération doublement indicée pour mettre en évidence la dépendance vis à vis du paramètre de régularisation :

$$x_k^{t+1} = ((\nabla h + \lambda_k B)^{-1} \circ \nabla h)((\nabla h + \lambda_k A)^{-1} \circ \nabla h)(x_k^t) \tag{6.12}$$

où $\{\lambda_k\}$ est une suite de scalaires strictement positifs.

De nouveau, il s'agit d'une méthode en deux temps. D'abord nous itérons sur l'indice t pour un k fixé. Ensuite nous augmentons k en générant un nouveau cycle d'itérations.

Dans le paragraphe suivant, nous appliquons notre technique de régularisation à la programmation convexe.

6.2.2 Application à la programmation convexe

Soit f une fonction définie de $\mathbb{R}^n$ dans $\mathbb{R} \cup \{+\infty\}$, propre, convexe et s.c.i.; et X un sous-ensemble convexe fermé, non vide de $\mathbb{R}^n$.
Nous savons que le problème de minimisation convexe

$$\text{(P1)} \qquad \begin{cases} \text{minimiser} & f(x) \\ \text{s.c.} & x \in X \end{cases}$$

est équivalent au problème :

$$\text{(P2)} \qquad \text{trouver } x \in \mathbb{R}^n \text{ tel que } 0 \in \partial(f + \delta_X)x$$

Pour les étapes intermédiaires, nous renvoyons le lecteur au paragraphe 2.1.

De plus, sous l'hypothèse $ri(\text{dom } \ f) \cap ri(X) \neq \emptyset$, ce problème peut se réécrire :

$$\text{(P3)} \qquad \text{trouver } x \in \mathbb{R}^n \text{ tel que } 0 \in \partial f(x) + \partial \delta_X(x)$$

où les opérateurs ∂f et $\partial(\delta_X(.))$ sont maximaux monotones puisque les fonctions f et δ_X sont propres, convexes et semi-continues inférieurement (cfr exemple 2.5).

Régularisons le problème (P3) en approximant l'opérateur $A := \partial f$.

L'itération (6.12) est alors la suivante :

$$x_k^{t+1} = ((\nabla h + \lambda_k \partial(\delta_X))^{-1} \circ \nabla h)((\nabla h + \lambda_k \partial f)^{-1} \circ \nabla h)(x_k^t) \tag{6.13}$$

où $\{\lambda_k\}$ est une suite de scalaires strictement positifs.

Fixons l'indice k et itérons sur l'indice t.

- D'une part, nous savons que

$$z_k^{t+1} = ((\nabla h + \lambda_k \partial f)^{-1} \circ \nabla h)(x_k^t)$$

  est l'unique élément de l'ensemble $\text{argmin}_{z \in \mathbb{R}^n} \{f(z) + \frac{1}{\lambda k} D_h(z, x_k^t)\}$ (cfr paragraphe 3.2).

- D'autre part l'opérateur $(\nabla h + \lambda_k \partial \delta_X)^{-1} \circ \nabla h$ est l'opérateur de D_h projection sur l'ensemble convexe fermé X (cfr proposition 2.7).

Dès lors pour k fixé, l'itération (6.13) se réécrit

$$x_k^{t+1/2} = \text{argmin}_{x \in I\!R^n} \{f(x) + \frac{1}{\lambda_k} D_h(x, x_k^t)\}$$

et

$$x_k^{t+1} \text{ est la } D_h \text{ projection de } x_k^{t+1/2} \text{ sur } X \ .$$

Ainsi *une itération proximale non linéaire sur la fonction f est couplée à la D_h projection sur l'ensemble convexe, fermé X.*

6.2.3 Convergence vers une solution des problèmes régularisés $(\mathbf{P}_{h\lambda})$

Nous considérons ici la première étape de notre méthode, étape où le paramètre de régularisation est considéré comme fixé dans les réels positifs.
Si nous posons $\lambda = \lambda_k$ et si nous omettons la dépendance de la suite $\{x^t\}$ vis à vis du paramètre λ, l'itération (6.12) se réécrit :

$$x^{t+1} = ((\nabla h + \lambda B)^{-1} \circ \nabla h)((\nabla h + \lambda A)^{-1} \circ \nabla h)(x^t)$$

Malheureusement, nous ne sommes pas parvenus à prouver la convergence de la suite $\{x^t\}$ vers un zéro de l'opérateur $A_{h\lambda} + B$.

En effet, nous avons d'abord essayé de suivre le schéma classique. Cependant il ne nous a pas été possible de montrer que les opérateurs $(\nabla h + \lambda A)^{-1} \circ \nabla h$ et
$(\nabla h + \lambda B)^{-1} \circ \nabla h$ sont des contractions.

Nous avons ensuite tenté d'imiter l'analyse de convergence de l'algorithme du point proximal non linéaire utilisant des fonctions de Bregman (Théorème 2.7). Mais nos efforts furent vains car nous n'avons pas pu vérifier la propriété suivante :

$$\text{si } z \in \overline{S} \text{ est un zéro de l'opérateur } A_{h_{lambda}} + B,$$
$$\text{et } P_y = ((\nabla h + \lambda B)^{-1} \circ \nabla h)((\nabla h + \lambda A)^{-1} \circ \nabla h)(y) \ .$$

Alors

$$D_h(z, P_u) \leq D_h(z, y) - D_h(P_y, y) \ .$$

En conclusion, nous pensons que les difficultés de généralisation sont dues à l'indépendance de la fonction de Bregman h et des opérateurs maximaux monotones A et B.

Chapitre 7

Conclusion

Dans ce mémoire, nous avons étudié une généralisation non linéaire de l'algorithme du point proximal de Rockafellar.

Après avoir analysé les propriétés d'existence et de convergence de cet algorithme nous l'avons appliqué aux formulations "primal", "dual" et "primal-dual" d'un problème de minimisation convexe.

L'application au "primal" a permis de définir une méthode de minimisation proximale non linéaire. Implémentée sur des architectures de calcul en parallèle, cette méthode peut résoudre un problème de programmation linéaire plus rapidement que les méthodes directes comme l'algorithme du simplexe ou les méthodes de points intérieurs.

L'application au "dual" et au "primal-dual" a donné naissance à une nouvelle classe de méthodes non quadratiques des multiplicateurs.

Cependant, plusieurs questions restent en suspens. Notamment :

- La convergence de l'algorithme du point proximal non linéaire sous une hypothèse plus faible que $\overline{D(T)} \subseteq S$.

- L'application de l'algorithme du point proximal non linéaire à la recherche d'un zéro d'une somme d'opérateurs maximaux monotones. Dans le chapitre 6, nous nous sommes intéressés à ce problème. Malheureusement, nous n'avons pas pu concrétiser nos recherches.

- La possibilité de calculer approximativement la suite des itérés. Ceci peut influencer l'implémentation des méthodes d'optimisation des chapitres 3, 4 et 5.

ANNEXE

Nous rassemblons ci-dessous quelques définitions et propriétés fondamentales de l'analyse convexe.

I. Ensembles et fonctions convexes

Soient C un sous-ensemble de $\mathbb{R}^n$ et f une fonction définie de $\mathbb{R}^n$ dans $\mathbb{R} \cup \{-\infty, +\infty\}$.

I.1. L'ensemble C est *convexe* si $\forall\ x, y \in C$ et $\forall\ \lambda \in [0,1]$, on a

$$\lambda x + (1-\lambda) y \in C\ .$$

I.2. Le *domaine* de la fonction f est l'ensemble

$$\text{dom}\ \ f := \{x \in \mathbb{R}^n \text{ tel que } f(x) < +\infty\}\ .$$

I.3. L'*épigraphe* de la fonction f est l'ensemble

$$epi\ f := \{(x, \mu) \in \mathbb{R}^n \times \mathbb{R} \text{ tel que } f(x) \leq \mu\}\ .$$

I.4. La fonction f est *propre* si

$$\text{dom}\ \ f \neq \emptyset \text{ et } f(x) > -\infty \quad \text{pour tout } x \in \mathbb{R}^n\ .$$

I.5. La fonction f est *convexe* si son épigraphe est un sous-ensemble convexe de $\mathbb{R}^{n+1}$.

I.6. Proposition : si la fonction f est propre, alors f est convexe si et seulement si dom f est convexe et $\forall\ x, y \in$dom f, $\forall\ \lambda \in [0,1]$

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)\ .$$

I.7. Si la fonction f est propre, alors f est *strictement convexe* si et seulement si dom f est convexe et $\forall\ x, y \in$dom f tel que $x \neq y$, $\forall\ \lambda \in]0,1[$

$$f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)\ .$$

I.8. Si la fonction f est propre, alors f *est fortement convexe* de rapport $\alpha (\alpha > 0)$ si dom f est convexe et $\forall\, x, y \in$dom f, $\forall\, \lambda \in]0,1[$

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) + \frac{1}{2}\alpha\lambda(1-\lambda)||x-y||^2 \; .$$

Enonçons maintenant quelques propriétés des fonctions convexes.

I.9. Proposition :

a) Si f est une fonction propre convexe et λ un scalaire positif alors la fonction λf est convexe.

b) Si f_1 et f_2 sont deux fonctions propres et convexes, alors la fonction $f_1 + f_2$ est convexe.

Preuve : cfr [19], théorème 5.2, p. 33.

I.10. Proposition : Soient (α, β) un intervalle ouvert de $I\!R$ et f une fonction définie de (α, β) dans $I\!R$ deux fois continûment différentiable.
Alors f est convexe, sur (α, β) si et seulement si sa dérivée seconde f'' est positive sur (α, β).

Preuve : cfr [19], théorème 4.4, p. 26.

I.11. Proposition : Soient C un sous-ensemble convexe ouvert de $I\!R^n$, et f une fonction définie de C dans $I\!R$ deux fois continûment différentiable.
Alors f est convexe sur C si et seulement si la matrice Hessienne

$$Q_x = (q_{ij}(x)) \text{ où } q_{ij}(x) = \frac{\partial^2 f}{\partial \xi_i \partial \xi_j}(\xi_1, \cdots, \xi_n)$$

est semi-définie positive pour tout $x \in C$.

Preuve : cfr [19], théorème 4.5, p. 27.

I.12. Proposition : Soit f une fonction convexe définie de $I \subseteq I\!R$ dans $I\!R$. Alors

$$\frac{f(x)-f(a)}{x-a} \leq \frac{f(b)-f(a)}{b-a} \leq \frac{f(b)-f(x)}{b-x}$$

pour tout $a, b, x \in I$ tel que $0 < x < b$.
De plus, si f est strictement convexe alors il s'agit d'inégalités au sens strict.

Preuve : cfr [21], théorème 1.4, p. 2.

I.13. Exemple : Soit C un ensemble convexe non vide de $I\!R^n$. Alors la fonction indicatrice de C, c'est-à-dire la fonction définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$ par

$$\delta_C(x) = \begin{cases} = 0 & \text{si } x \in C \\ +\infty & \text{sinon} \end{cases}$$

est convexe.

II. Intérieur relatif

Soit A un sous-ensemble de $I\!R^n$.

II.1. La *variété affine* de A est le plus petit sous-ensemble affine de $I\!R^n$ qui contient A. Elle est notée aff (A).

II.2. L'*intérieur relatif de A, noté* $ri(A)$, est l'intérieur de A par rapport à aff (A). Autrement dit,

$$ri(A) := \{x \in \text{ aff } (A) \text{ tel que } \exists \varepsilon > 0, (x + \varepsilon B) \cap (\text{aff } (A)) \subset A\}$$

où B désigne la boule unité de $I\!R^n$.

II.3. L'ensemble A est *relativement ouvert* si $ri(A) = A$.

Enonçons maintenant quelques propriétés de l'intérieur relatif d'un ensemble convexe.

II.4. Proposition : Soit C un sous-ensemble convexe non vide de $I\!R^n$. Alors $riC \neq \emptyset$.

Preuve : cfr [19], théorème 6.2, p. 45.

II.5. Proposition : Soit C un sous-ensemble convexe de $I\!R^n$. Alors $\overline{ri(C)} = \overline{C}$ et $ri(\overline{C}) = ri(C)$.

Preuve : cfr [19], théorème 6.3, p. 46.

II.6. Proposition : Soit C un sous-ensemble convexe de $I\!R^n$. Alors tout sous-ensemble ouvert qui coupe $\overline{C}$, coupe aussi $ri(C)$.

Preuve : cfr [19], corollaire 6.3.2, p. 46.

II.7. Proposition : Soient C et D deux sous-ensembles convexes de $I\!R^n$ et λ un réel. Alors

(i) $ri(\lambda C) = \lambda ri(C)$

(ii) $ri(C + D) = ri(C) + ri(D)$
Si de plus $ri(C) \cap ri(D) \neq \emptyset$ alors

(iii) $\overline{C \cap D} = \overline{C} \cap \overline{D}$

(iv) $ri(C \cap D) = ri(C) \cap ri(D)$.

Preuve : cfr [21], théorème 4.10, p. 47.

II.8. Proposition : Soient D un ensemble ouvert de $I\!R^n$ et C un ensemble convexe tels que $D \subset C$. Alors $D \subset ri(C)$.

Preuve :
Puisque $D \subset C \subset \overline{C}$, nous déduisons de la proposition A.II.6. que $D \cap riC \neq \emptyset$.
Or $ri(D) = D$ car D est un ouvert, d'où par la proposition A.II.7 (iv) :

$$ri(D \cap C) = riD \cap riC\ .$$

Par conséquent, comme $D \subset C$ et $ri(D) = D$, nous obtenons que $D = D \cap ri(C)$, c'est-à-dire $D \subset ri(C)$. ∎

II.9. Proposition : Soit C un ensemble convexe de $I\!R^n$. Alors $\overline{C}$ et riC sont des ensembles convexes de $I\!R^n$ engendrant la même variété affine. Leur dimension est donc égale à celle de C.

Preuve : cfr [9], théorème 6.2, p. 45.

III. Semi-continuité inférieure et fermeture

III.1. La fonction f définie de $\mathbb{R}^n$ dans $\mathbb{R} \cup \{-\infty, +\infty\}$ est *semi-continue inférieurement* (ou s.c.i.) en $x_0 \in \mathbb{R}^n$ si

$$\forall\, \lambda < f(x_0)\ \exists \delta > 0 \ \text{ tel que } \ \forall\, x \in \mathbb{R}^n \ \ ||x - x_0|| < \delta \Rightarrow \lambda < f(x)$$

III.2. Proposition : La fonction f définie de $\mathbb{R}^n$ dans $\mathbb{R} \cup \{-\infty, +\infty\}$ est semi-continue inférieurement en x_0 si et seulement si

$$f(x_0) = \lim_{x \to x_0} \ \inf \ f(x)$$

III.3. On définit *la fermeture* de la fonction convexe propre $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$ par $cl\ f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$ où

$$(cl\ f)(x) = \sup_{\substack{F \text{ affine} \\ F \leq f}} \ F(x)$$

Nous remplacerons dorénavant $(cl\ f)(x)$ par $cl\ f(x)$ (abus de notation).

III.4. Proposition : Soit f une fonction définie de $\mathbb{R}^n$ dans $\mathbb{R} \cup \{+\infty\}$ convexe et propre. Alors on les quatre propriétés suivantes sont équivalentes

(i) f est semi-continue inférieurement sur $\mathbb{R}^n$

(ii) *epi* f est fermé dans $\mathbb{R}^{n+1}$

(iii) $\forall\, a \in \mathbb{R}$, l'ensemble de niveau $\{x \text{ tel que } f(x) \leq a\}$ est fermé dans $\mathbb{R}^n$.

(iv) $cl\ f = f$ c'est-à-dire f est fermée.

Preuve : cfr [19], théorème 7.1, p. 51.

III.5. Proposition : Soient f une fonction propre, convexe, fermée, définie de $\mathbb{R}^n$ dans $\mathbb{R} \cup \{+\infty\}$, et y un vecteur de $\mathbb{R}^n$.
S'il existe $x \in$dom f tel que $\lim_{\lambda \to +\infty}$ inf $f(x + \lambda y) < +\infty$
alors $\lim_{\lambda \to +\infty}$ inf $f(x + \lambda y) < +\infty$ pour tout $x \in$dom f.

III.6. Exemple : Soit X un ensemble convexe, fermé, non vide de $\mathbb{R}^n$ et δ_X la fonction indicatrice de cet ensemble, c'est-à-dire

$$\delta_X(x) = \begin{cases} 0 & \text{si } x \in X \\ +\infty & \text{sinon.} \end{cases}$$

Alors δ_X est une fonction propre, convexe et semi-continue inférieurement sur $\mathbb{R}^n$.

IV. Dérivée directionnelle et sous-gradients

A. Dérivée directionnelle

Soit f une fonction définie de $\mathbb{R}^n$ dans $\mathbb{R} \cup \{-\infty, +\infty\}$, et $x \in \mathbb{R}^n$ tel que $f(x)$ est fini.

IV.1. Pour tout vecteur y de $\mathbb{R}^n$, la *dérivée directionnelle* de f en x dans la direction y est la limite

$$f'(x;y) = \lim_{\lambda \downarrow 0} \frac{f(x+\lambda y) - f(x)}{\lambda}$$

si cette limite existe ($+\infty$ et $-\infty$ sont des valeurs permises).

IV.2. : Proposition : si f est différentiable en x, alors les dérivées directionnelles $f'(x;y)$ sont finies pour tout $y \in \mathbb{R}^n$.
De plus, $f'(x;y) =< \nabla f(x), y >$, où $\nabla f(x)$ est le gradient de f en x.

IV.3. Proposition : Si f est une fonction convexe, alors pour tout

$$y, f'(x;y) \text{ existe et } f'(x;y) = \inf_{\lambda>0} \frac{f(x+\lambda y) - f(x)}{\lambda}$$

Preuve : cfr [19], théorème 23, p. 213.

B. Sous-gradients d'une fonction convexe

Considérons dorénavant une fonction convexe f définie de $\mathbb{R}^n$ dans $\mathbb{R} \cup \{-\infty, +\infty\}$.

IV.4. Soit $x \in \mathbb{R}^n$.
Un vecteur $x^* \in \mathbb{R}^n$ est un *sous-gradient* de f au point x si

$$f(z) \geq f(x) + < x^*, z - x1 \ \forall z \, .$$

IV.5. L'ensemble des sous-gradients de f en x est appelé le *sous-différentiel de f en x*. Il est noté $\partial f(x)$.

IV.6. Proposition : $\partial f(x)$ est un ensemble convexe et fermé pour tout $x \in \mathbb{R}^n$. De plus l'ensemble $\partial f(x)$ est vide pour tout $x \notin \text{dom } f$.

IV.7. Le *sous-différentiel de f* est l'opérateur multivoque défini de $\mathbb{R}^n$ dans l'ensemble des parties de $\mathbb{R}^n$ par $\partial f : x \to \partial f(x)$.
De plus f est dite *sous-différentiable* en x si $\partial f(x) \neq \emptyset$.

IV.8. Le domaine effectif et l'image du sous-différentiel de f sont respectivement définis par :

$$D(\partial f) := \{x \in \mathbb{R}^n \text{ tel que } \partial f(x) \neq \emptyset$$
$$\text{et } R(\partial f) := U\{\partial f(x) \text{ tel que } x \in \mathbb{R}^n\}$$

IV.9. <u>Proposition</u> : Nous avons que

$$ri(\text{dom } \ f) \subset D(\partial f) \subset \text{dom } \ f$$
$$\text{et } \ ri(\text{dom } \ f^*) \subset R(\partial f) \subset \text{dom } \ f^*$$

où la fonction f^* est la conjuguée de Fenchel de f.

Les deux théorèmes ci-dessous établissent un lien entre le sous-différentiel de f en x et la dérivée directionnelle de f en x.

IV.10. <u>Proposition</u> : Soient f une fonction convexe et $x \in \mathbb{R}^n$ tel que $f(x)$ est fini. Alors x^* est un sous-gradient de f en x si et seulement si $f'(x;y) \geq < x^*, y > \ \forall\, y$.

<u>Preuve :</u> cfr [19], théorème 23.2, p. 216.

IV.11. <u>Proposition</u> : Soient f une fonction propre, convexe et $x \in ri(\text{dom } \ f)$. Alors $\partial f(x)$ est non vide, et $f'(x;y)$ est une fonction propre et fermée en y.
De plus $f'(x;y) = \sup\{< x^*, y > \text{ tel que } x^* \in \partial f(x)\}$.

<u>Preuve :</u> cfr [19], théorème 23.4, p. 217.

Enonçons maintenant quelques propriétés utiles pour le calcul des sous-gradients.

IV.12. <u>Proposition</u> : [sous-différentiel d'une somme]
Soient $f_1 + \cdots + f_m$ des fonctions propres, convexes sur $\mathbb{R}^n$, et $f := f_1, \cdots, f_m$. Alors

$$\partial f(x) \supseteq \partial f_1(x) + \cdots + \partial f_m(x) \quad \forall\, x\,.$$

De plus si $\cap_{i=1}^m ri(\text{dom } \ f_i) \neq \emptyset$ alors

$$\partial f(x) = \partial f_1(x) + \cdots + \partial f_m(x) \quad \forall\, x\,.$$

<u>Preuve :</u> cfr [19], théorème 23.8, p. 224.

IV.13. Proposition : Soit $f(x) = h(Ax)$ où h est une fonction propre, convexe sur $I\!R^m$ et A une transformation linéaire de $I\!R^n$ dans $I\!R^m$. Alors $\partial f(x) \supseteq A^* \partial h(x) \quad \forall\ x$.
De plus si (image A) $\cap ri(\text{dom} \ \ h) \neq \emptyset$,
ou si h est polyhédrale et image $A) \cap (\text{dom} \ \ h) \neq \emptyset$
alors

$$\partial f(x) = A^* \partial h(Ax) \quad \forall\ x\ .$$

Preuve : cfr [19], théorème 23.9, p. 225.

IV.14. Proposition : Soit f une fonction propre et convexe.
Alors

$$\partial f(x) = \partial f(x) + N_{\text{dom } f}(x) \text{ pour tout } x \in I\!R^n$$

Preuve :
Soit $x \in I\!R^n$ arbitraire.

- Montrons d'abord que $\partial f(x) + N_{\text{dom } f}(x) \subseteq \partial f(x)$.
  En effet, $\forall\ u \in \partial f(x)$, $\forall\ v \in N_{\text{dom } f}(x)$ nous savons que

$$f(z) \geq f(x) + < u, z - x > \ \backslash; \ \forall\ z \in \text{dom} \ \ f \tag{1}$$

  et

$$< v, z - x > \leq 0 \quad \forall\ z \in \text{dom} \ \ f \tag{2}$$

  Sommant (1) et (2), nous obtenons que

$$f(z) \geq f(x) + < u + v, z - x > \quad \forall\ z \in \text{dom} \ \ f$$

  c'est-à-dire $u + v \in \partial f(x)$.
- D'autre part, il est évident que $\partial f(x) \subseteq \partial f(x) + N_{\text{dom } f}(x)$.
  En effet,

$$\forall\ y \in \partial f(x), \ \ y + 0 \in \partial f(x) + N_{\text{dom } f}(x)$$

- Par conséquent, il résulte que $\partial f(x) = \partial f(x) + N_{\text{dom } f}(x)$. ■

IV.15. Proposition : Soient $f_i \ i = 1 \cdots m$ des fonctions propres, convexes, définies de $I\!R^n$ dans $I\!R \cup \{+\infty\}$, et y un vecteur de $I\!R^n$ tel que $y \geq 0$.
Supposons également que $\cap_{i=1}^m ri(\text{dom} \ \ f_i) \neq \emptyset$. Définissons la fonction L de $I\!R^n$ dans $I\!R \cup \{+\infty\}$ par $L(x) = \sum_{i=1}^m y_i f_i(x)$.
De plus nous prenons comme convention que

$$y_i f_i(x) = +\infty \quad \text{si } y_i = 0 \text{ et } f_i(x) = +\infty$$

Alors pour tout $x^0 \in \cap_{i=1}^m \text{dom } f_i$,

$$\partial L(x^0) = \sum_{i=1}^m y_i \partial f_i(x^0) + N_{(\cap_{i=1}^m \text{dom } f_i)}(x^0) .$$

Preuve :

- Puisque par hypothèse les fonctions $y_i f_i$ $i = 1, \cdots, m$ sont convexes et $\cap_{i=1}^m ri(\text{dom } y_i f_i) = \cap_{i=1}^m ri(\text{dom } f_i) \neq \emptyset$, nous déduisons du théorème concernant le sous-différentiel d'une somme (proposition A.IV.12) que

$$\partial L(x^0) = \sum_{i=1}^m \partial(y_i f_i)(x^0) \tag{3}$$

- Evaluons maintenant $\partial(y_i f_i)(x^0)$ pour i arbitraire.
  Par convention, la fonction $y_i f_i$ est définie de $\mathbb{R}^n$ dans $\mathbb{R} \cup \{+\infty\}$ par

$$(y_i f_i)(x) = \begin{cases} y_i f_i(x) & \text{si } x \in \text{dom } f_i \\ +\infty & \text{si } x \notin \text{dom } f_i \end{cases}$$

  - Si $y_i = 0$ alors $0 f_i = \delta_{(\text{dom } f_i)}$, ainsi $\partial(0 f_i)(x^0) = N_{(\text{dom } f_i)}(x^0)$.
  - Si $y_i > 0$ alors $y_i f_i = y_i f_i + \delta_{(\text{dom } f_i)}$, et donc

$$\partial(y_i f_i)(x^0) = y_i \partial f_i(x^0) + N_{(\text{dom } f_i)}(x^0)$$

    D'autre part, il suit d'une propriété du sous-différentiel que $\partial(y_i f_i)(x^0) = y_i \partial f_i(x^0)$. C'est pourquoi $N_{(\text{dom } f_i)}(x^0)$ peut être négligé si $y_i > 0$.

  Par conséquent, nous obtenons que

$$\partial(y_i f_i)(x^0) = y_i \partial f_i(x^0) + N_{(\text{dom } f_i)}(x^0)$$

  où $N_{(\text{dom } f_i)}(x^0)$ peut être négligé si $y_i > 0$.

- Remplaçant $\partial(y_i f_i)$ par sa valeur, nous déduisons de (3) que

$$\partial L(x^0) = \sum_{i=1}^m y_i \partial f_i(x^0) + \sum_{i=1}^m N_{(\text{dom } f_i)}(x^0)$$

  De plus, puisque $\cap_{i=1}^m ri(\text{dom } f_i) \neq \emptyset$, il découle de [corollaire 23.8] que

$$\partial L(x^0) = \sum_{i=1}^m y_i \partial f_i(x^0) + N_{(\cap_{i=1}^m \text{dom } f_i)}(x^0) .$$

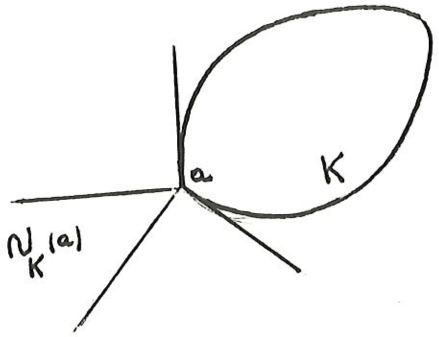
∎

V. Cône normal

Soit K un ensemble non vide de $I\!R^n$ et $a \in K$

V.1. Le *cône normal* à K en a est l'ensemble défini par

$$N_K(a) := \{x^* \in I\!R^n \text{ tel que } < x - a, x^* > \leq 0 \ \forall\, x \in K\}$$



V.2. Proposition : pour tout $x \in K$ $N_K(x) = \partial\delta_K(x)$ où $\delta_K(.)$ est la fonction indicatrice de l'ensemble K.

Preuve :
Par définition

$$x^* \in \partial\delta_K(x) \Leftrightarrow \delta_K(z) \geq \delta_K(x) + < x^*, z - x > \quad \forall\, z \in I\!R^n$$

Ainsi, comme $x \in K$ nous obtenons que

$$x^* \in \partial\delta_K(x) \Leftrightarrow \delta_K(z) \geq 0 + < x^*, z - x > \quad \forall\, z \in I\!R^n$$

c'est-à-dire

$$x^* \in \partial\delta_K(x) \Leftrightarrow 0 \geq < x^*, z - x > \quad \forall\, z \in K\ .$$

■

V.3. Proposition : Soient A et B deux ensembles convexes de $I\!R^n$ tels que $A \subseteq B$. Alors $N_A(x) + N_B(x) = N_A(x)\ \forall\, x \in B$.

Preuve : Soit un vecteur $x \in B$ arbitraire.

- Montrons d'abord que $N_A(x) + N_B(x) \subseteq N_A(x)$.

  En effet, $\forall\, x_1 \in N_A(x)$ et $\forall\, x_2 \in N_B(x)$, nous savons que

$$< x_1, z - x > \leq 0 \;\; \forall\, z \in A \subseteq B \tag{4}$$

  et

$$< x_2, z - x > \leq 0 \;\; \forall\, z \in B \tag{5}$$

  Sommant (4) et (5), nous obtenons que

$$< x_1 + x_2, z - x > \leq 0 \;\; \forall\, z \in A\,.$$

  c'est-à-dire $x_1 + x_2 \in N_A(x)$.

- D'autre part, il est évident que $N_A(x) \subseteq N_A(x) + N_B(x)$ car $\forall\, y \in N_A(x) \quad u + 0 \in N_A(x) + N_B(x)$

- Par conséquent, nous obtenons que $N_A(x) + N_B(x) = N_A(x)$. ∎

VI. La conjuguée de Fenchel

VI.1. Soit f une fonction convexe définie de $\mathbb{R}^n$ dans $\mathbb{R} \cup \{-\infty, +\infty\}$. Alors la fonction f^* définie de $\mathbb{R}^n$ dans $\mathbb{R} \cup \{-\infty, +\infty\}$ par

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \{< x, y > -f(x)\}$$

est appelée la fonction *conjuguée de Fenchel* de f.

VI.2. La fonction conjuguée de Fenchel de la fonction indicatrice δ_X d'un ensemble convexe, fermé X est appelée *fonction d'appui*. Elle est caractérisée par

$$\delta_X^*(y) = \sup_{x \in X} < x, y > \;.$$

VI.3. Exemples :

1. Si $f(x) = \frac{1}{2}||x||^2$ alors $f^*(x^*) = \frac{1}{2}||x^*||^2$ où $||.||$ désigne la norme euclidienne sur $I\!R^n$.
2. Si $f(x) = e^x$ où $x \in I\!R$ alors

$$f^*(x^*) = \begin{cases} x^* \ln x^* - x^* & \text{si } x^* > 0 \\ 0 & \text{si } x^* = 0 \\ +\infty & \text{si } x^* < 0 \end{cases}$$

De plus $f^{**} = f$.

Preuve : cfr [19], pp. 105-106.

Enonçons maintenant deux propriétés importantes de la conjuguée de Fenchel.

VI.4. Proposition : Si la fonction f est convexe et propre, alors la fonction conjuguée de f, ie f^*, est propre, convexe et fermée. De plus $(clf)^* = f^*$ et $f^{**} = clf$ où clf désigne la fermeture de la fonction f.

Preuve : cfr [19], théorème 12.2.

VI.5. Proposition : Si la fonction f est propre, convexe et fermée, alors ∂f^* est l'inverse de ∂f dans le sens des opérateurs multivoques, c'est-à-dire

$$x \in \partial f^*(x^*) \Leftrightarrow x^* \in \partial f(x) .$$

Preuve : cfr [19], corollaire 23.5, p. 219.

Pour terminer, montrons que l'addition et la convolution infimale sont des opérations duales.

VI.6. Soient f et g deux fonctions propres et convexes sur $I\!R^n$. Alors la *convolution infimale* de f et g est définie par $(f\Box g)(x) = \inf_y \{f(x-y)+g(y)\}$.
De plus dom $(f\Box g) =$ dom $f+$ dom g.

VI.7. Soient $f_1, \cdots, f_m$ des fonctions propres et convexes sur $I\!R^n$. Alors

$$(f_1\Box\cdots\Box f_m)^* = f_1^* + \cdots + f_m^*$$

et

$$(clf_1 + \cdots + cff_m)^* = cl(f_1^*\Box\cdots\Box f_m^*)$$

De plus si $\cap_{i=1}^m ri(\text{dom } f_i) \neq \emptyset$ alors

$$(f_1 + \cdots + f_m)^* = f_1^*\Box\cdots\Box f_m^* .$$

<u>Preuve :</u> cfr [19], théorème 16.4, p. 145.

VII. Propriétés extrémales des fonctions convexes

Considérons le problème suivant

$$\begin{cases} \text{Minimiser} & f(x) \\ \text{s.c.} & x \in I\!R^n \end{cases}$$

où f est une fonction propre, convexe, définie de $I\!R^n$ dans $I\!R \cup \{+\infty\}$.

VII.1. <u>Proposition</u> : soit $x_0 \in I\!R^n$.
x^0 est un minimum de f sur $I\!R^n$ si et seulement si $0 \in \partial f(x^0)$.

VII.2. <u>Proposition</u> : si la fonction f est strictement convexe, alors l'ensemble des minima de f sur $I\!R^n$, c'est-à-dire argmin $_{x\in I\!R^n} \{f(x)\}$, est soit l'ensemble vide, soit un singleton.

VII.3. <u>Proposition</u> : Soit f une fonction propre, convexe et fermée. Alors
argmin $_{x\in I\!R^n} \{f(x)\} = \{\overline{x}\} \Leftrightarrow f^*$ est différentaible en 0 et $\overline{x} = \nabla f^*(0)$.

<u>Preuve :</u> cfr [19], théorème 27.1, p. 265.

VIII. Cone et fonction de récession

La notion de cone de récession permet d'étudier les ensembles convexes, fermés et non bornés de $I\!R^n$.

VIII.1. Soit C un ensemble convexe non vide de $I\!R^n$. Le *cone de récession* de C est l'ensemble des vecteurs y de $I\!R^n$ tels que

$$x + \lambda y \in C \quad \forall\, \lambda \geq 0 \text{ et } \forall\, x \in C$$

Il est noté 0^+C.

Notons qu'un vecteur y non nul de 0^+C est appelé une *direction de récession* de C.

Le thorème suivant indique qu'un ensemble convexe fermé non borné contient au moins une *direction de récession*, c'est-à-dire au moins un point à l'infini.

VIII.2. Proposition : Soit C un ensembe convexe, fermé, non vide de $I\!R^n$. C est borné si et seulement si $0^+C = \{0\}$ où 0 est le vecteur nul.

Appliquons maintenant la notion de récession aux fonctions convexes.

Soit f une fonction convexe définie sur $I\!R^n$. Supposons qu'il existe un vecteur x de $I\!R^n$ tel que $f(x) \neq +\infty$.
Par définition, $(y, v) \in 0^+$ (epi f) si et seulement si

$$(x, u) + \lambda(y, v) = (x + \lambda y, u + \lambda v) \in epi\ f$$

pour tout couple $(x, u) \in epi\ f$ et pour tout $\lambda \geq 0$.
Autrement dit, $f(x + \lambda y) \leq f(x) + \lambda v \;\; \forall\, x$ et $\forall\, \lambda \geq 0$.

Par conséquent, pour un y donné, les valeurs de v pour lesquelles $(y, v) \in 0^+(epi\ f)$ forment un intervalle fermé, non borné supérieurement de $I\!R$, ou l'intervalle vide. Ainsi le cone de récession $0^+(epi\ f)$ est le graphe d'une certaine fonction.

VIII.3. La *fonction de récession* de f, notée $f0^+$, est la fonction dont l'épigraphe est le cone de récession de $epi\ f$, c'est-à-dire $epi(f0^+) = 0^+(epi\ f)$.

Pour terminer, énonçons quelques propriétés des fonctions de récession.

VIII.4. Proposition : Soit f une fonction propre, convexe, fermée sur $I\!R^n$. Alors la fonction de récession de f est une fonction propre, convexe, fermée, et positivement homogène.

De plus, pour tout $x \in$dom $f, f0^+$ est défini par la formule suivante :

$$(f0^+)(y) = \sup_{\lambda>0} \frac{f(x+\lambda y) - f(x)}{\lambda} = \lim_{\lambda\to\infty} \frac{f(x+\lambda y) - f(x)}{\lambda}$$

Preuve : cfr [19], théorème 8.5, p. 66.

VIII.5. Proposition : Soient f une fonction propre, convexe et y un vecteur de $I\!R^n$. Supposons également qu'il existe $x \in$dom f tel que$\lim_{\lambda\to+\infty}$ inf $f(x+\lambda y) < +\infty$. Alors $\lim_{\lambda\to+\infty}$ inf $f(x+\lambda y) < +\infty$ pour tout x si et seulement si $(f0^+)y \leq 0$.

Preuve : cfr [19], théorème 8.6, p. 68.

VIII.6. Proposition : Soient f_1 et f_2 deux fonctions propres, convexes et fermées sur $I\!R^n$, telles que

$$(f_1 0^+)(z) + (f_2 0^+)(-z) > 0 \text{ pour tout } z \neq 0$$

Alors la convolution infimale de f_1 et f_2, c'est-à-dire $f_1 \Box f_2$, est une fonction propre, convexe et fermée.

De plus, pour tout x, l'infimum de l'expression suivante

$$(f_1 \Box f_2)(x) = \inf_y \{f_1(x-y) + f_2(y)\}$$

est atteint par un certain vecteur $\overline{y}$.

Preuve : cfr[19], corollaire 9.2.1, p. 76.

TABLE DES MATIERES

Bibliographie

[1] D.P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press New-York, 1982.

[2] H. Brézis. *Opérateurs Maximaux Monotones et Semi-groupes de Contractions dans les Espaces de Hilbert.* Mathematics Studies 5, North-Holland, 1973.

[3] H. Brézis and A. Haraux. "Image d'une somme d'opérateurs monotones et applications". *Israel J. Math*, **23(2)**, (1976), pp. 165-186.

[4] Y. Censor and A. Lent. "An iterative row-action method for interval convex programming". *J. Optim. Theory Applic.*, **34(3)**, (1981), pp. 321-353.

[5] Y. Censor and J. Segman. "On block-Iterative Entropy Maximization". *Journal of Information and Optimization Sciences*, **Vol. 8**, (1987), pp. 275-291.

[6] Y. Censor and S.A. Zenios. *The proximal minimization algorithm with D-functions.* Report 89-12-17, Decision Sciences Department, Wharton School, University of Pennsylvania, Philadelphia, PA, 1990.

[7] Y. Censor and S.A. Zenios. *On the use of D-function in primal-dual methods and in the proximal minimization algorithm.* Report 91-04-03, Decision Sciences Department, Wharton School, University of Pennsylvania, Philadelphia, PA, 1990.

[8] C. Ghen and M. Teboulle. *Convergence Analysis of a Proximal-Like Minimization Algorithm Using Bregman Functions.* Research Report 90, 23, Department of Mathematics and Statistics, UMBC, 1990.

[9] G.B. Dantzig. *Linear Programming and Extensions.* Princeton University Press, Princeton, New Jersey, 1963.

[10] J. Eckstein. Non Linear Proximal Point Algorithms Using Bregman Functions, with Applications To Convex Programming. Working Paper 91-004, Harvard Business School, 1990.

[11] J.R. Eriksson. *An Iterative Primal-Dual Algorithm for Linear Programming.* Report LITH-MAT-R-1985-10, Department of Mathematics, Linköping University, Linköping, Sweden.

[12] N. Karmarkar. "A new polynomial algorithm for linear programming", *Combinatorica* **4**, (1984), pp. 373-395.

[13] F.J. Luque. *The Nonlinear Proximal Point Algorithm.* Report LIDS-P-1598, Laboratory for Information and Decision Sciences, Massachusetts Institute of Technology, Cambridge, MA, 1986.

[14] P. Mahey and Pham Dinh Tao. "Partial Regularization of the sum of two maximal monotone operators". Working paper, Artemis/Imag, 1990.

[15] B. Martinet. "Régularisation d'inéquations variationnelles par approximations successives". Rairo 4, (1970), pp. 154-159.

[16] K. Mouallif. *Convergence variationnelle et méthodes perturbées pour les problèmes d'optimisation et de point-selle.* Thèse d'Etat, Université de Liège 1989.

[17] R.T. Rockafellar. "Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming". *Math. Oper. Res.,* **1(2)**, (1976), pp. 97-116.

[18] R.T. Rockafellar. "A Dual Approach to solving Nonlinear Programming Problems by Unconstrained Optimization". *Mathematical Programming* **5**, (1973), pp. 354-373.

[19] R.T. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[20] R.T. Rockafellar. "Monotone Operators and the Proximal Point Algorithm", **SIAM** *Journal Control and Optimization,* **14(5)**, (1976), pp. 877-898.

[21] J. Van Tiel. *Convex Analysis, An Introductory Text.* John Willey and Sons LTD, 1984.

[22] F. Warnier. *Méthode de projection proximale et techniques de décomposition en programmation convexe.* Mémoire, Facultés Universitaires Notre-Dame de la Paix, Namur, 1991.